Myosin heavy chain, a major component of the contractile apparatus, is a ubiquitously protein found in muscle and non-muscle tissues throughout the animal kingdom (1-5). Multiple MHC isoforms have been identified by immunological (6, 7), and peptide analyses (7-9). The various muscle types express structurally distinct MHC cDNA clones (17-19). It has been determined previously that the MHC mRNAs expressed in striated muscle tissues exhibit sequence homology within an organism as well as throughout metazoan evolution (20). Thus, MHC is encoded by a multigene family, similar to several other genes coding for contractile and noncontractile proteins (8, 21-24). A comparative study of the structural organization of the individual members of the MHC multigene family would provide insight into the relationship between gene structure and the mechanisms regulating the differential expression of the various MHC isoforms. In this study, we have used pMHC-25, a recombinant clone containing MHC structural gene sequences from a rat myogenic cell line (17), to isolate MHC genomic sequences from a rat spleen genomic library. These clones have been used to analyze the structural organization of some of the developmentally regulated MHC genes.

**EXPERIMENTAL PROCEDURES**

**Screening of Phage Genomic Libraries**—The isolation of MHC genomic sequences was performed under the conditions described by Blattner et al. (25). Approximately 1.5 \times 10^{10} phage plaques, from a partial EcoRI genomic library of rat genomic spleen DNA, were screened using 32P-labeled pMHC-25 DNA, obtained by the nick-translation procedure (26).

**DNA Blot Hybridization**—70 pg of each genomic clone and 10 pg of genomic DNA L6E9 or rat kidney were restricted with EcoRI as described by suppliers (Bethesda Research Laboratories). The EcoRI-digested DNAs were size-fractionated in 0.7470 agarose gels in 0.2% sodium dodecyl sulfate at 55 °C. Ethidium bromide at 50 V for 18 h. The size-fractionated DNAs were denatured, transferred to nitrocellulose filters (27), and hybridized to 32P-labeled pMHC-25 DNA. After hybridization, the filters were washed, with several changes, in 1 \times SSC at 37 °C. Kodak XAR-5 film was exposed to nitrocellulose filter for 4 days at -70 °C with one lightning Plus-X intensifying screen.

**Preparation of RNA**—Total cytoplasmic RNA was isolated from the L6E9 cells and rat fibroblast as described (28). Total RNA was extracted from the striated muscle tissue by the hot phenol procedure (29). To eliminate DNA, the RNA was first precipitated with guanidine hydrochloride followed by ethanol precipitation. Poly(A+) RNA was purified from total RNA by affinity chromatography using oligo(dT)-cellulose (Collaborative Research) (30).

**RNA Blot Hybridization**—10 pg of total RNA was size-fractionated in 1% agarose, 3% formaldehyde gels in 20 mm 4-morpholinepropanesulfonic acid (Sigma), pH 7.4, 1 mM EDTA at 200 V for 3 h. Transfer to nitrocellulose filters (Millipore) and hybridization to 32P-labeled plasmid probes were done as described (20). At end of hybridization, the filters were washed, with several changes, in 0.1 \times SSC, 0.2% sodium dodecyl sulfate at 35 °C.

**Heteroduplex Analysis**—For each pair-wise heteroduplex combination, phage particles of each clone, representing approximately 0.25 pg of DNA, were mixed and denatured in 0.1 M NaOH, 10 mm EDTA for 10 min at 37 °C. Renaturation was done in 50% formamide at room temperature for 30 min. Aliquots were spread on 10% formamide hypophase as described (31).

**R-loop Analysis**—R-loop formation and purification were performed as described (32) using 0.5 pg of DNA from the genomic clone 287A-3 or 287A-4 and 15 pg of poly(A+) RNA from L6E9 myotubes or adult skeletal muscle, respectively. Subsequent formamide spreading was done according to the procedure of Chow and Broker (33).

**RESULTS**

**Isolation of MHC Gene Sequences**—A 625-bp cDNA clone, pMHC-25, constructed from RNA isolated from the myogenic cell line L6E9 (17), contains MHC mRNA sequences corresponding to the rat embryonic skeletal muscle MHC. This recombinant plasmid shows substantial hybridization to all sarcomeric MHC mRNAs tested (20). However, pMHC-25...
does not hybridize, even under low stringency conditions, to nonsarcomeric MHC mRNA. This observation, which is not due to differences in MHC mRNA concentration (18, 20), demonstrates that sarcomeric MHC mRNAs have at least partial sequence homology which is not shared by the nonsarcomeric MHC mRNAs. Thus, pMHC-25 serves as a suitable probe to screen for sarcomeric MHC gene sequences.

The isolation of the MHC genomic sequences was carried out essentially as described (25). Approximately \(1.5 \times 10^6\) phage plaques, from a partial EcoRI genomic library of rat genomic DNA, were screened using \(^{32}\)P-labeled pMHC-25 DNA as the probe. From this screening, six clones were isolated. Preliminary restriction maps showed that two of these clones, although the result of independent cloning events, contain inserts which are a subset of a third. Four different genomic clones were selected for further study. Each clone contains from 3 to 7 EcoRI restriction fragments (Fig. 1A) and the size of the insert ranges from 11 to 16 kb. Clones 287A-3 and 287A-6 have some similarities. The 2.5-kb EcoRI restriction fragment in both clones (Fig. 1A), pA3-B and pA6-B, respectively (see expanded maps in Fig. 1B), have identical restriction maps. This similarity extends to part of the DNA sequence to the right of these fragments. Despite these similarities, clones 287A-3 and 287A-6 do not represent overlapping fragments of the same gene but two different gene sequences. This is demonstrated by the different restriction sites to the right of the XbaI site in the pA3-F and pA6-A fragments (Fig. 1B). The similarities between these two clones will be discussed below. Based on the restriction maps in Fig. 1B, the genomic clones represent four unique, nonoverlapping sequences.

pMHC-25, which does not contain any EcoRI restriction sites (17), hybridizes to a minimum of eight bands in EcoRI-digested LEr DNA (see Fig. 2 and Ref. 20). To determine how many of these bands are represented in the genomic clones, EcoRI digests of DNA isolated from the four clones, LEr myogenic cells, and rat kidney were size-fractionated, transferred to nitrocellulose paper, and hybridized to \(^{32}\)P-labeled pMHC-25 DNA. As shown in Fig. 2 (lane 3), pMHC-25 hybridizes to the EcoRI fragment of clone 287A-4 corresponding to the fragment designated band 5 in both LEr cell and rat genomic DNAs (lanes 5 and 6). Clones 287A-3 and 287A-6 (lanes 2 and 4) each have two EcoRI fragments which hybridize to pMHC-25. One of these bands corresponds to the 2.5-kb EcoRI fragment which is common to the two clones and is represented by band 7 in the genomic DNAs (lanes 5 and 6). The second hybridizing band in clone 287A-3 (lane 2) corresponds to band 8 in the LEr and rat DNAs. The second hybridizing band in clone 287A-6 (lane 4), however, is not detected in LEr or any tissue DNAs, including sperm, examined so far (data not shown). The results from restriction mapping (Fig. 1B) and hybridization studies (Fig. 2) suggest that clone 287A-6 represents an allelic form of the MHC gene corresponding to clone 287A-3. Since clones 287A-3 and 287A-

![Ethidium bromide-stained gel of EcoRI-restricted MHC genomic clones (A) and partial restriction maps of the MHC genomic clones (B).](image-url)
Characterization of Sarcomeric Myosin Heavy Chain Genes

672 Characterization of Sarcomeric Myosin Heavy Chain Genes

The sources of DNA were: denatured, transferred 287A-4; clone 287A-1; 5 were isolated from the same recombinant genomic library, to the band designated The pMHC-25 hybridizing band in clone 287A-1 corresponds gene sequence. This possibility is presently being investigated. from a polymorphic strain and heterozygous for this MHC myotubes; fragment which occurs in varying frequencies in different rat populations (20). Thus, the results described above show that there is another sarcomeric MHC gene family which is to multiple bands in the rat genome (Fig. 2, lane 6). We have shown previously that the genomic fragments labeled 3 and 3' (lanes 5 and 6, respectively) represent allelic forms of a MHC gene fragment which occurs in varying frequencies in different rat populations (20). Thus, the results described above show that four of the pMHC-25-hybridizing bands in the L6& and rat genomic DNAs are represented in these four genomic clones. The existence of multiple MHC genes, as detected by pMHC-25, is consistent with published data on the MHC proteins. A minimum of nine different sarcomeric MHC proteins have been identified (7,10,11). This number is probably an underestimate, because some of the MHC proteins appear to be very similar (18) and would not be detected by the biochemical techniques used so far to analyze the MHC proteins (1). However, only seven sarcomeric MHC gene sequences can be detected by pMHC-25 on genomic DNA blots (Fig. 2, lane 5). This indicates that individual bands in the genomic blot may represent more than one MHC gene or that there is another sarcomeric MHC gene family which is not detected by pMHC-25. The results presented here do not allow a distinction between these two possibilities.

Tissue Specificity of MHC Gene Sequences—It has been established previously that pMHC-25 contains embryonic skeletal muscle MHC mRNA sequences (20) and hybridizes to multiple bands in the rat genome (Fig. 2, lane 5). All four MHC genomic clones isolated with this probe could represent embryonic skeletal muscle MHC gene sequences. Alternatively, each genomic clone could represent a different MHC gene sequence, expressed in a particular muscle tissue or stage of development. To distinguish between these two possibilities, each genomic clone and their respective subcloned EcoRI fragments (see Fig. 1B for subclone designations) were 32P-labeled and hybridized to Northern blots of RNAs, isolated from various rat tissues. As shown in Fig. 3 (A, C, D, and G), each clone preferentially hybridizes to MHC mRNA from a particular striated muscle or stage of development. This result indicates that each genomic clone contains a different MHC gene sequence. The pattern of hybridization is, however, not completely specific. All four clones hybridize to RNA from more than one tissue and, upon extended exposure, hybridize to MHC mRNA from all striated muscle tissues. No hybridization to MHC mRNA from smooth muscle, nonmuscle (fibroblasts), and undifferentiated L6& myoblasts was observed (Fig. 3F). This absence of hybridization is not due to different concentrations of MHC sequences in these RNA preparations since the same results were obtained when poly(A+) RNA was used (data not shown). These results indicate either that each sarcomeric MHC gene sequence is expressed at different levels in each striated muscle type or that there exist varying degrees of sequence homology among the sarcomeric MHC mRNAs. This ambiguity is resolved by the hybridization patterns obtained with the subcloned EcoRI fragments (see Fig. 1B for subclone designations).

Genomic clone 287A-1 hybridizes most strongly to MHC mRNA from adult cardiac muscle and weakly to MHC mRNA from adult skeletal muscle and differentiated L6& myotubes (Fig. 3A). This pattern of hybridization, however, is not uniform throughout the MHC insert. Subclone pA1-A, which contains the pMHC-25 hybridizing region, and subclone pA1-B show strong hybridization to MHC mRNA from adult skeletal muscle and cardiac muscle of newborn rats (Fig. 3B). In contrast, subclone pA1-C hybridizes exclusively to MHC mRNA from cardiac muscle. It is, therefore, most likely that clone 287A-1 contains a gene sequence coding for an adult cardiac MHC. This is further reinforced by the observation that clone 287A-1 hybridizes strongly to cardiac MHC cDNA clones isolated from adult ventricular muscle (18).

The genomic clone 287A-4 hybridizes selectively to adult skeletal muscle RNA (Fig. 3C). Data obtained with the subclones are not presented. Since the adult skeletal muscle RNA used consists of a mixture of slow and fast muscle types (1, 10), it is not possible to determine whether clone 287A-4 contains slow or fast skeletal muscle MHC gene sequences. However, a fast skeletal muscle MHC cDNA clone which has been isolated in our laboratory hybridizes strongly to clone 287A-4.

The genomic clones 287A-6 and 287A-3 have similar hybridization patterns to RNA from various muscle tissues, as expected from their restriction maps (Fig. 1B) and DNA blot hybridizations to pMHC-25 (Fig. 2). Genomic clone 287A-6 (Fig. 3G) shows preferential, but not exclusive, hybridization to MHC mRNA from L6& myotubes (embryonic skeletal muscle MHC). Although pMHC-25 hybridizes to two EcoRI fragments in this clone (Fig. 2), one of them, pA6-A, shows selective hybridization to MHC mRNA from L6& myotubes. In contrast, the other fragment, pA6-B, shows additional strong hybridization to MHC mRNA from adult and newborn skeletal muscle (Fig. 3H). Similarly, genomic clone 287A-3 shows preferential hybridization to MHC mRNA from L6& myotubes (Fig. 3D). Of the two EcoRI fragments that hybridize to pMHC-25, pA3-E is selective for MHC mRNA from L6& myotubes and skeletal muscle of newborn rats (Fig. 3F) while pA3-B shows hybridization to MHC mRNA from several other striated muscle tissues. These results strongly suggest that clones 287A-3 and 287A-6 contain gene sequences coding for an embryonic form of skeletal muscle MHC. Since the hybridization to fetal RNA is weak and the hybridization to the RNA from the L6& myotubes intense, it is probable that there may be more than one MHC gene which is characteristic of the prenatal stage. Recently, Whalen et al. (7) have shown data supporting the existence of a neonatal skeletal muscle MHC which appears prior to birth and persists throughout adolescence in rats. According to these authors, the parental cell line L6 (from which the L6& cells were

2 D. H. Hornig, and B. Nadal-Ginard, unpublished results.
The tissue specificity of the genomic clones and their EcoRI fragments has been confirmed by washing the blots shown in Fig. 3 at different stringencies. In general, the tissue specificity of each clone becomes more evident at high stringency while it is more difficult to discern at low stringency. However, the different degrees of cross-hybridization of different subclones, as shown in Fig. 3, are still evident after the high stringency washes. The results clearly indicate that each of the genomic clones contains an EcoRI fragment that is specific for a particular tissue and/or developmental stage as well as fragments that hybridize to MHC mRNA from several muscle tissues. This demonstrates an interesting feature of the MHC gene, namely the existence of highly conserved regions, common to several MHC genes, that are located among gene-specific sequences. This conclusion is further reinforced by the observation that each EcoRI fragment of the genomic clones, when used to probe a Southern blot of rat genomic DNA, hybridizes only to the EcoRI band that corresponds to the size of the probe. However, after long exposure, the pMHC-25 hybridizing fragments (Fig. 2) hybridize weakly to all the pMHC-25 hybridizing bands seen in genomic DNA.

**Heteroduplex Analysis of MHC Gene Sequences**—Through DNA sequence analysis, we have determined that pMHC-25 codes for a portion of the light meromyosin near the carboxyl terminus of MHC (34, 35). Since the four genomic MHC clones hybridize to pMHC-25, they must also contain sequences coding for light meromyosin. As shown above, these sequences have been highly conserved in different striated muscle tissues and developmental stages. They have also been conserved in a variety of different eukaryotic species (20). To determine the organization of the conserved sequences and the gene-specific regions as well as their location in this region of the light meromyosin sequence, the extent of homology between the genomic sequences in the different clones was determined by DNA heteroduplex mapping. From direct sequence analysis of genomic clone 287A-3 and comparison with the pMHC-25 sequence, the 5’ to 3’ orientation of clone 287A-3 with respect to the MHC mRNA was determined. The orientation of clone 287A-6 could be inferred from the similarities of its restriction map with clone 287A-3 (see Fig. 1B). The DNA inserts of clones 287A-3 and 287A-6 are reversed with respect to the 5’ to 3’ orientation of genomic clones 287A-1 and 287A-4 was determined by their ability to form heteroduplexes with clone 287A-6. The restriction maps of the genomic clones in Fig. 1B are shown in the 5’ to 3’ (left to right) orientation.

The abbreviations correspond to the following rat cells and tissues from which the RNAs were obtained. Sm.M., smooth muscle from uterine tissue; Sk.M., hind leg adult skeletal muscle; L.Eo, Mb, differentiated L.Eo myotubes; L.Eo, Mb, undifferentiated L.Eo myoblasts; H, adult cardiac muscle; GH, rat pituitary cell line (nonmuscle); Fib, primary fibroblasts (nonmuscle); Ft.M., pooled fetal skeletal muscle from 16-day-old gestation period; Nb M, skeletal muscle from newborn (1-day-old); Nb H, cardiac muscle from newborn (1-day-old); Fet us, pooled fetal muscle from 16-day-old gestation period. A, hybridization with the genomic clone 287A-1. B, hybridization with the EcoRI subclones of the genomic clone 287A-1 (see Fig. 1B for subclone designations). C, hybridization with the genomic clone 287A-4. D, hybridization with the EcoRI subclones of the genomic clone 287A-3. E, a longer exposure of the blot shown in D. F, hybridization with the EcoRI subclones of the genomic clone 287A-3. G, hybridization with the genomic clone 287A-6. Hybridization with the EcoRI subclones of the genomic clone 287A-6.

**Fig. 3.** Tissue specificity of the MHC genomic clones and the EcoRI subclones. 10 μg of total RNA from rat myogenic and nonmyogenic sources were size-fractionated in 1% agarose, 3% formaldehyde denaturing gels, transferred to nitrocellulose filters, and hybridized to 32P-labeled MHC genomic clones or EcoRI subclones. The abbreviations correspond to the following rat cells and tissues from which the RNAs were obtained. Sm.M., smooth muscle from uterine tissue; Sk.M., hind leg adult skeletal muscle; L.Eo, Mb, differentiated L.Eo myotubes; L.Eo, Mb, undifferentiated L.Eo myoblasts; H, adult cardiac muscle; GH, rat pituitary cell line (nonmuscle); Fib, primary fibroblasts (nonmuscle); Ft.M., pooled fetal skeletal muscle from 16-day-old gestation period; Nb M, skeletal muscle from newborn (1-day-old); Nb H, cardiac muscle from newborn (1-day-old); Fet us, pooled fetal muscle from 16-day-old gestation period. A, hybridization with the genomic clone 287A-1. B, hybridization with the EcoRI subclones of the genomic clone 287A-1 (see Fig. 1B for subclone designations). C, hybridization with the genomic clone 287A-4. D, hybridization with the genomic clone 287A-3. E, a longer exposure of the blot shown in D. F, hybridization with the EcoRI subclones of the genomic clone 287A-3. G, hybridization with the genomic clone 287A-6. Hybridization with the EcoRI subclones of the genomic clone 287A-6.

3. M. Periasamy, R. M. Wydro, H. T. Nguyen, and B. Nadal-Ginard, manuscript in preparation.
Approximately 15 heteroduplexed molecules were analyzed for each pair-wise combination. Typical heteroduplexed molecules of the possible combinations are shown in Fig. 4, A–C (heteroduplexes involving genomic clone 287A-3 were not possible since the insert in this clone is in the opposite 5' to 3' orientation with respect to the other genomic clones). In all of the heteroduplex combinations analyzed, there is only a small region of homology (900–1500 bp in length) centering around the sequences that hybridize to pMHC-25. This homologous region is interrupted by two small regions of non-homology. In order to analyze the distribution of these homologous sequences, the regions of duplex formation were mapped onto the restriction maps of each genomic clone. Several interesting features of the MHC genes emerge from the summary drawing in Fig. 4D. Although each possible pair-wise combination shows three short regions of homology interrupted by two short non-homologous regions, the size and distribution of both the homologous and non-homologous sequences seem to be different for each combination. Precise measurement of DNA molecules longer than 30 kbp is difficult. However, it appears that, although there are regions common among all three clones (as would be expected since all three were isolated through hybridization to pMHC-25), some non-homologous regions in one combination are homologous in another. This result does not suggest that homologous regions among the different clones represent exon sequences interrupted by introns of different size in each genomic sequence. This organization of the genomic MHC sequences, which needs to be confirmed by nucleotide sequence analysis, could represent a combination of different intron lengths and varying degree of sequence conservation among the different MHC genes.

Fig. 4D shows that clone 287A-4 contains more sequence to the left (towards the 5' end of the gene) than either clone 287A-1 or 287A-6. The 3' end of the MHC sequence contained in clone 287A-1 was mapped by hybridization to the 3' untranslated region of a cardiac MHC cDNA clone (18), as indicated under the line drawing of 287A-1 (data not shown). This indicates that the remaining DNA to the right of this region represents flanking DNA sequences which are not expected to be homologous among different MHC genes. The alignment of the clones in Fig. 4B suggests that clone 287A-6 would also contain the 3' end of its respective gene. Indeed, recent DNA sequencing analysis of this clone in our laboratory has shown that the 3' end of the gene is at a similar location as that in clone 287A-1 (data not shown).

It is important to note the apparent discrepancy between the RNA blot data presented in Fig. 3 and the data presented in Fig. 4D. The data in Fig. 4D indicate that most of the DNA which is not involved in duplex formation between clones

![Fig. 4. Heteroduplex analyses of the MHC genomic clones.](image)
287A-6 and 287A-1 is 3' flanking DNA (outside the gene coding for MHC). This is consistent with the lack of hybridization of subclone pA6-C to MHC mRNA from any muscle tissue analyzed so far (data not shown). However, the subcloned fragments pA1-B and pA1-D (see Figs. 1B and 3B) hybridize specifically to MHC mRNA from adult cardiac muscle. This result strongly suggests that the flanking sequences in clone 287A-1 are part of an adjacent MHC gene coding for an adult cardiac MHC mRNA. Since this region does not hybridize to the 3' untranslated region of a cardiac MHC cDNA clone (18), this MHC genomic sequence most probably represents the 5' end of an adjacent MHC gene. Whether this sequence is part of a functional MHC gene or a pseudogene is currently under investigation.

R-loop Analysis of MHC Genomic Clones 287A-3 and 287A-4—The results from the RNA blots and heteroduplex analysis suggest that the MHC genes have an unusual gene organization. We have demonstrated recently, by DNA sequence analysis and RNA blots probed with a subfragment of clone 287A-3 corresponding to an intervening sequence, that the MHC gene represented by clone 287A-3 is expressed during LEm myogenesis. This cell line thus provides a good source of homogeneous MHC mRNA with which to determine the distribution of RNA coding sequences along this gene. Towards this end, we analyzed R-loops formed through hybridization of clone 287A-3 to poly(A') RNA from LEm myotubes. A representative of the resulting hybrid molecules (Fig. 5A) shows a minimum of 18 loops representing intervening sequences (see Fig. 5B for line drawing). Several molecules were photographed and measured to produce the consensus molecule depicted in Fig. 5E. The exons in the segment of the MHC gene represented by clone 287A-3 are short with an average length of approximately 200 bp. The total exon length in clone 287A-3 is approximately 4650 bp and accounts for approximately 65% of the full length MHC mRNA. From the protein data available (35), this region of the genomic clone codes for most of the S2 fragment, the complete light meromyosin region, and the carboxyl terminus of the MHC molecule. The introns show considerable variation in size ranging from approximately 100 to 1000 bp. The intron:exon ratio for this portion of the gene is 1.6:1. If this ratio were maintained for the remainder of the sequence, this embryonic skeletal muscle MHC gene, coding for a 7100-nucleotide mRNA (17), is approximately 18.5 bp in length and contains approximately 30 intervening sequences. This organization of the MHC gene represents one of the most complex examples of gene structure reported so far. We have isolated recently the 5' end and flanking sequence of the MHC gene represented by 287A-3. The analysis of these sequences is in progress.

R-loop analysis was also performed with genomic clone 287A-4. Although formal proof of its expression is lacking, the data presented in Fig. 3C suggest that it is expressed in adult rat skeletal muscle. We analyzed R-loops formed through hybridization of clone 287A-4 to poly(A') RNA from adult rat skeletal muscle. A representative molecule is shown in Fig. 5c with its accompanying line drawing (Fig. 5D). A diagrammatic representation, the result of several molecules, is shown in Fig. 5E. A minimum of 11 intervening sequences can be identified in this genomic clone. The exons in the segment of the MHC gene represented by clone 287A-4 are short with an average length of approximately 200 bp. The intervening sequences have considerable variability in size, ranging from 200 to 1000 bp. The intron:exon ratio for this portion of the gene is 1.2:1. If this ratio were maintained for the remainder of the sequence, this embryonic skeletal muscle MHC gene, coding for a 7100-nucleotide mRNA (17), is approximately 18.5 bp in length and contains approximately 30 intervening sequences. This organization of the MHC gene represents one of the most complex examples of gene structure reported so far. We have isolated recently the 5' end and flanking sequence of the MHC gene represented by 287A-3. The analysis of these sequences is in progress.

R-loop analysis was also performed with genomic clone 287A-4. Although formal proof of its expression is lacking, the data presented in Fig. 3C suggest that it is expressed in adult rat skeletal muscle. We analyzed R-loops formed through hybridization of clone 287A-4 to poly(A') RNA from adult rat skeletal muscle. A representative molecule is shown in Fig. 5c with its accompanying line drawing (Fig. 5D). A diagrammatic representation, the result of several molecules, is shown in Fig. 5E. A minimum of 11 intervening sequences can be identified in this genomic clone. The exons in the segment of the MHC gene represented by clone 287A-4 are short with an average length of approximately 200 bp. The intervening sequences have considerable variability in size, ranging from 200 to 1000 bp. The intron:exon ratio for this portion of the gene is 1.2:1. If this ratio were maintained for the remainder of the sequence, this embryonic skeletal muscle MHC gene, coding for a 7100-nucleotide mRNA (17), is approximately 18.5 bp in length and contains approximately 30 intervening sequences. This organization of the MHC gene represents one of the most complex examples of gene structure reported so far. We have isolated recently the 5' end and flanking sequence of the MHC gene represented by 287A-3. The analysis of these sequences is in progress.
100 bp to greater than 1000 bp. The total exon length in clone 287A-4 is approximately 6400 bp and accounts for approximately 88% of the MHC mRNA. The region of the MHC protein molecule coded in this segment of the gene is not known. However, the heteroduplex analysis and hybridization data with cloned adult skeletal muscle cDNAs suggest that this clone does not contain the 3' end of the mRNA although it contains sequences coding for the carboxyl terminus of the MHC protein (data not shown). Although no heteroduplexes could be formed between clones 287A-3 and 287A-4, it is possible to align the R-loops of the two molecules due to the 2.5-kb EcoRI fragment shared between clones 287A-3 and 287A-6 (see Figs. 1C and 4D). With the possible exception of the last exon at the 3' end, it is evident that both clones correspond approximately to the same region of the MHC gene, starting at or very near the 3' end of the gene, and that clone 287A-4 contains more coding sequences than clone 287A-3. Clone 287A-4 is interrupted by only 13 introns while the shorter clone 287A-3 is interrupted by 18. Additionally, the differences between the two clones is evident throughout the length of the molecules, both in the size and distribution of exons, being most obvious at the 3' end. These results suggest that the number, size, and distribution of exons have not been conserved between embryonic (clone 287A-3) and adult (clone 287A-4) skeletal MHC genes. Confirmation of this result will be obtained through nucleotide sequencing analyses.

DISCUSSION

In this report, we have presented data on the isolation and characterization of several genomic DNA fragments containing rat sarcomeric MHC gene sequences. Each of the clones appears to be expressed in a tissue-specific manner and is developmentally regulated. The organization of these MHC genomic sequences is unusual. The coding regions of the individual members of this multigene family share small regions of sequence homology flanked by gene-specific regions of divergent DNA sequence. In addition, the coding region of at least two of these genes is interrupted by multiple intervening sequences, whose number, length as well as distribution of some of them do not appear to be conserved in the two genes.

In studying a multigene family, several questions must be addressed. First, how many members make up the multigene family? Second, what is the structure of the individual member genes and how are they related to each other? Third, is the expression of members of the multigene family coordinately or independently regulated?

The precise number of sarcomeric MHC genes in the rat genome has not yet been determined. Published data and results from our own laboratory suggest that there are a minimum of 11 sarcomeric MHC genes (18-20). This number is probably an underestimate. A minimum of nine sarcomeric MHC proteins have been identified (7, 10, 11). Recently, a large number of myosin forms, differing in electrophoretic mobility on pyrophosphate gels, have been shown to exist in fast and slow skeletal muscle (24). Whether these electrophoretic forms are due to differences in the embryonic and the putative adult rat skeletal MHC gene. However, due to the limitations of the R-loop and heteroduplex analyses, the existence of introns less than 50 bp would not have been detected. Nonetheless, recent DNA sequence analysis of clone 287A-3 has not uncovered any new introns in the 5 kb located at the 3' end of the gene. In contrast to the two rat MHC genes analyzed so far, the unc-54 body wall MHC in Caenorhabditis elegans contains only seven introns while a Drosophila MHC gene has only four introns. Moreover, the positions of most introns in these four MHC genes do not appear to be conserved. Clearly, more data on the MHC genes within the same and from other species are necessary before conclusions on the significance of this organization can be drawn.

The third intriguing aspect of multigene families is the expression of the individual member genes and the importance of gene organization in their expression. The histone multigene family in the sea urchin shows a coordinate expression of the repeating gene cluster during embryogenesis (47). The individual members of the B-like globin (48) and immunoglobulin (39) gene families are expressed sequentially in a 5' to 3' progression during development and immune response. The actin gene family, although dispersed throughout the genome, by gene-specific sequences, both at the gene and mRNA level. This organization is in contrast to the data available on the B-like globin (36), vitellogenin (37), and actin (38) multigene families which exhibit considerable conservation of most of the coding sequences. The sequence organization of the MHC genes is somewhat similar to the organization of the immunoglobulin genes where the constant and variable regions allow for the maintenance of common functional domains in molecules that exhibit antigenic specificity (39). In the MHC molecule, the necessity to evolve different contractile properties while maintaining the ability of the MHC to assemble into the thick filament could explain the different rates at which various regions of the molecule have diverged.

The biological significance of intervening sequences has not yet been determined. It has been postulated that exons define structural or functional domains of proteins based on the structure of the immunoglobulin, insulin, and globin genes (40). The actin and collagen gene families appear to be exceptions to this generalization. Dicystostelum discoideum (21) and some Drosophila melanogaster (38) actin genes do not contain any intervening sequences, while sea urchin (41), yeast (42), and other Drosophila (38) actin genes contain a single intron (43). Moreover, some vertebrate actin genes contain more than one intervening sequence (26). Additionally, the intron location, is different in each of these species. Moreover, in Drosophila actin genes, the intron when present is found in at least three different locations (38). There is no evidence that the different locations where introns are found separate functional domain on the actin molecule. Similarly, the Drosophila collagen gene contains only one intron (44), while its vertebrate counterparts are interrupted by many (45).

An analogous situation appears to be prevalent in the MHC multigene family. The R-loop analyses of the embryonic and the adult skeletal muscle MHC genes presented here show that the gene sequence coding for the S2 fragment and light meromyosin region of the MHC protein is interrupted by multiple intervening sequences. Both the light meromyosin and S2 fragments of the MHC protein have a uniform coiled-coil structure (46); thus, it is unlikely that the multiple introns separate functional domains of the MHC protein. Furthermore, there appear to be striking differences in the number, size and distribution of introns between the embryonic and the putative adult rat skeletal MHC gene. However, due to the limitations of the R-loop and heteroduplex analyses, the existence of introns less than 50 bp would not have been detected. Nonetheless, recent DNA sequence analysis of clone 287A-3 has not uncovered any new introns in the 5 kb located at the 3' end of the gene. In contrast to the two rat MHC genes analyzed so far, the unc-54 body wall MHC in Caenorhabditis elegans contains only seven introns while a Drosophila MHC gene has only four introns. Moreover, the positions of most introns in these four MHC genes do not appear to be conserved. Clearly, more data on the MHC genes within the same and from other species are necessary before conclusions on the significance of this organization can be drawn.

The third intriguing aspect of multigene families is the expression of the individual member genes and the importance of gene organization in their expression. The histone multigene family in the sea urchin shows a coordinate expression of the repeating gene cluster during embryogenesis (47). The individual members of the B-like globin (48) and immunoglobulin (39) gene families are expressed sequentially in a 5' to 3' progression during development and immune response. The actin gene family, although dispersed throughout the genome,
is also regulated temporally and spatially regulated (49). A single α-actin gene (actin 1) appears, however, to be expressed in all Drosophila muscle types (50). The MHC genes appear to be more complex. The sequence divergence among different types of MHC genes appears to be greater than among the actin genes, where the divergence mainly in the noncoding region of the genes (51). Thus far, we have been unable to detect nonsarcomeric (smooth muscle and cytoplasmic) MHC gene sequences by hybridization to cDNA and genomic sarcomeric MHC clones. Biochemical and immunological data suggest that the nonsarcomeric MHCs are distinct from either vertebrate actins, the sarcomeric and nonsarcomeric MHCs by a multigene family that is regulated temporally and spatially. Similar to the globins

Acknowledgments—We thank Drs. Gregory Goldberg and Frederick Blattner for help in isolation of the genomic clones, Dr. T. Broker, Dr. L. Chow, and Jane Fant for help with the electron microscopy, Eva Bekesi for excellent technical assistance, and Dr. Russell Medford for critical reading of the manuscript.

REFERENCES

1. Brevet, A., and Whalen, R. G. (1978) Biochimie 60, 459-466
2. Rubin, N. A., Pepe, F. A., and Haltzer, H. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 4924-4927
3. Sartore, S., Pierobon-Bornioli, S., and Schiaffino, S. (1978) Nature 274, 82-83
4. Pollard, T. D., and Weighing, R. (1974) CRC Crit. Rev. Biochem. 2, 1-65
5. Clarke, M., and Spudich, I. A. (1977) Annu. Rev. Biochem. 46, 797-822
6. Masaki, T., and Kinosita, T. (1974) J. Biochem. (Tokyo) 76, 123-131
7. Whalen, R. G., Sell, S. M., Butler-Browne, G. S., Schwartz, K., Bouveret, P., and Finset-Harström, I. (1981) Nature 292, 895-899
8. Whalen, R. G., and Sell, S. M. (1980) Nature 286, 731-733
9. Whalen, R. G., Schwartz, K., Bouveret, P., Sell, S. M., and Gros, F. (1979) Proc. Natl Acad. Sci. U. S. A. 76, 5197-5202
10. Gauthier, G. F., Lowrey, S., and Hobbs, A. W. (1978) Nature 274, 25-29
11. Hoh, J. F. Y., McGrath, P. A., and Hale, H. T. (1978) J. Mol. Cell. Cardiol. 10, 1053-1076
12. Whillingham, M. C., Omel, R. E., and Pastan, I. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 4144-4148
13. Wagner, P. D., and Giniger, E. (1981) Nature 252, 560-562
14. Streeter, F. A., Seidel, J. C., and Gergely, T. (1966) J. Biol. Chem. 241, 5772-5776
15. Flak, I., L., Rader, J. H., and Morkin, E. (1979) J. Biol. Chem. 254, 3103-3110
16. Lompre, A.-M., Schwartz, K., d’Albis, A., Lacombe, G., van Thiemen, N., and Swynghedauw, B. (1979) Nature 282, 105-107
17. Medford, R. M., Wydro, R. M., Nguyen, H. T., and Nadal-Ginard, B. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5749-5755
18. Mahdavi, V., Periasamy, M., and Nadal-Ginard, B. (1982) Nature 297, 659-664
19. Nudel, V., Katocoff, D., Carmon, Y., Zevin-Sonkin, D., Levi, Z., Shaul, Y., Shani, M., and Yaffe, D. (1980) Nucleic Acids Res. 8, 2123-2146
20. Nguyen, H. T., Gubits, R. M., Wydro, R. M., and Nadal-Ginard, B. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 8290-8294
21. Firtel, R. A. (1981) Cell 24, 67
22. Engel, J. N., Gunning, P. W., and Kedes, L. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 4674-4678
23. Shani, M., Zevin-Sonkin, D., Saxel, O., Carmon, Y., Katocoff, D., Nudel, V., and Yaffe, D. (1981) Dev. Biol. 86, 483-492
24. Hoh, J. F. Y., and Yeoh, G. P. S. (1980) Nature 280, 321-323
25. Blattner, F., Blech, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slighton, J. L., Tucker, P. W., and Smith, O. (1978) Science (Wash. D. C.) 202, 1279-1284
26. Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-255
27. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517
28. Benoff, S., and Nadal-Ginard, B. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 1853-1857
29. Sööro, R., Birnboim, H. C., and Darnell, J. E. (1966) J. Mol. Biol. 19, 362-372
30. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412
31. Ferguson, J., and Davia, R. W. (1978) in Advanced Techniques in Biological Electron Microscopy 2 (Koehler, J. K., ed) Springer-Verlag, New York
32. Woolford, J. L., and Roshbash, M. (1959) Nucleic Acids Res. 6, 2483-2497
33. Chow, L. T., and Broker, T. R. (1975) The Electron Microscopy of Nucleic Acids, Cold Spring Harbor Laboratory, New York
34. Lowrey, S., Sluyter, H. S., Weeds, A. G., and Baker, H. (1969) J. Mol. Biol. 42, 1-29
35. Elzenga, M., and Trus, B. (1980) Methods in Peptide and Protein Sequencing Analysis (Birr, C., ed) pp. 213-224, Elsevier, Amsterdam
36. Hardison, R. C., Butler, E. T., III, Lacy, E., Manisits, T., Rosenhal, N., and Eifrat, J. H. (1979) Cell 18, 1285-1297
37. Wall, W., David, I. B., Ryffel, G. U., and Weber, R. (1981) Science (Wash. D. C.) 212, 298-304
38. Fyrberg, E. A., Bond, B. J., Hensley, N. D., Mithers, K. S. and Davidson, N. (1981) Cell 24, 107-116
39. Davis, M. M., Kim, S. K., and Howard, L. (1980) Cell 22, 1-2
40. Gilbert, W. (1979) in Eukaryotic Gene Regulation (Axel, R., Manisits, T., and Fox, C. F., eds) pp. 1-12, Academic Press, New York
41. Schuler, M. A., and Keller, E. (1981) Nucleic Acids Res. 9, 391-404
42. Ng, R., and Abelson, J. (1980) Proc. Natl Acad. Sci. U. S. A. 77, 4674-4678
43. Ordahl, C. P., Tilghman, S. M., Ott, C., Forwald, J., and Largen, M. T. (1980) Nucleic Acids Res. 8, 4989-5006
44. Mack, J. M., Natla, J., Friedman, J., and McCarthy, B. J. (1983) Proc. Natl Acad. Sci. U. S. A. 79, 1761-1765
45. Woznoff, J., Hanahaa, D., Tate, V., Boedttery, H., and Dity, P. (1981) Nature 294, 129-135
46. Needs, A. G., and Pope, B. (1977) J. Mol. Biol. 111, 129-157
47. Houstschel, C. C., and Linsel, M. L. (1981) Cell 25, 301-313
48. Jeffrey, A. J., and Flavell, R. A. (1977) Cell 12, 429-439
49. Zulaff, E., Sanchez, F., Tchau, S. L., Rodest, V., and McCarthy, B. J. (1981) Nature 292, 556-558

* L. Leinwand and B. Nadal-Ginard, unpublished results.
Characterization of Sarcomeric Myosin Heavy Chain Genes

50. Storti, R. V., Horovitch, S. J., Scott, M. P., Rich, A., and Pardue, M. L. (1978) *Cell* 13, 589-598
51. Frederiksen, D. W. (1979) *Biochemistry* 18, 1651-1656
52. Burridge, K., and Bray, D. (1975) *J. Mol. Biol.* 99, 1-14
53. Weatherall, D. J., and Clegg, J. B. (1979) *Cell* 16, 467-479
54. Holt, J. F. Y., Kwan, B. T. S., Dunlop, C., and Kim, B. R. (1980) in *Plasticity of Muscle* (Pette, D., ed) Academic Press, New York