Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 3504–3508, May 1989
Biochemistry

Characterization of human cardiac myosin heavy chain genes

(gene organization/DNA sequence)

KEIKO YAMAUCHI-TAKIHARA, MICHAEL J. SOLE, JACK LIEW, DOUG ING, AND CHOONG-CHIN LIEW

Laboratory of Molecular Cardiology, Departments of Medicine, Physiology, and Clinical Biochemistry, Max Bell Research Center, Toronto General Hospital, University of Toronto, Toronto, ON, M5G 1L5, Canada.

Communicated by C. C. Tan, January 13, 1989 (received for review December 2, 1988)

ABSTRACT **We have isolated and analyzed the structure of the genes coding for the α and β forms of the human cardiac myosin heavy chain (MYHC). Detailed analysis of four overlapping MYHC genomic clones shows that the α-MYHC and β-MYHC genes constitute a total length of 51 kilobases and are tandemly linked. The β-MYHC-encoding gene, predominantly expressed in the normal human ventricle and also in slow-twitch skeletal muscle, is located 4.5 kilobases upstream of the α-MYHC-encoding gene, which is predominantly expressed in normal human atrium. We have determined the nucleotide sequences of the β form of the MYHC gene, which is 100% homologous to the cardiac MYHC cDNA clone (pHMC3). It is unlikely that the divergence of a few nucleotide sequences from the cardiac β-MYHC cDNA clone (pHMC3) reported in a MYHC cDNA clone (pSMHCZ) from skeletal muscle is due to a splicing mechanism. This finding suggests that the same β form of the cardiac MYHC gene is expressed in both ventricular and slow-twitch skeletal muscle. The promoter regions of both α- and β-MYHC genes, as well as the first four coding regions in the respective genes, have also been sequenced. The sequences in the 5′-flanking region of the α- and β-MYHC-encoding genes diverge extensively from one another, suggesting that expression of the α- and β-MYHC genes is independently regulated.**

Considerable evidence indicates that several muscle-specific contractile proteins, including the myosin heavy chain (MYHC), are constructed by multigene families. Each different sarcomeric MYHC gene displays a pattern of expression that is tissue specific and developmentally regulated (1–6). In mammalian myocardium, the MYHCs exist as dimers within the myosin molecule; two α-MYHCs form a homodimer called V1, one α- and one β-MYHC form a heterodimer called V2, and two β-MYHCs form a homodimer called V3 (7, 8). In small mammals, the alternative expression of ventricular α- and β-MYHCs is not only developmentally regulated but also governed by thyroxine (4, 9, 10) or increased hemodynamic load (11, 12).

So far, the ventricular α- and β-MYHC mRNA sequences and their corresponding genes have been isolated and characterized in the rat (3, 13) and rabbit (4, 14). Recently, human α- and β-MYHC mRNA sequences (15, 16) and their corresponding genes have been isolated and partially characterized (17–20).

We have isolated the complete human cardiac α- and β-MYHC genes from the genomic library. As the α- and β-MYHC genes are tandemly linked, we characterized and sequenced the genes* to provide the basis for further investigation into the regulatory mechanism for the expression of cardiac MYHC genes.

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.

MATERIALS AND METHODS

Materials. All chemicals are of reagent grade. Restriction enzymes were purchased from Pharmacia, Boehringer Mannheim, and Amersham. ^{32}P-labeled nucleotides were obtained from New England Nuclear.

Isolation and Characterization of Cardiac MYHC Genes. Human cosmid genomic libraries provided by Y. W. Kan (University of California, San Francisco) and L. Chan (Baylor Medical College, Houston) were screened with human cardiac β-MYHC cDNA clone (pHMC3) (15) used as probe. The human genomic library in bacteriophage λ Charon 4A constructed by Maniatis *et al.* (21) was screened with a 5′ DNA fragment of clone Cos HM-1 (pE3P1) used as probe, by following described methods (15). DNA fragments from the genomic clones were subcloned into either the M13mp18 or mp19 vectors, and DNA sequences were determined by the dideoxy chain-termination method or a specific primer-directed dideoxy chain-termination method (22).

Southern Blot Analysis. High-molecular-weight genomic DNA prepared from human peripheral blood leukocytes as described (23) was digested with restriction endonucleases *Eco*RI, *Bam*HI, and *Hin*dIII. These digested DNA fragments were electrophoresed on 0.8% agarose gel and transferred onto a nitrocellulose filter (BA85; Schleicher & Schuell). The blots were probed with a ^{32}P-labeled 1.8 kilobase (kb) *Sac* I fragment from clone Cos HM-1 (pE5S1).

RESULTS

Identification of α- and β-Forms of the Cardiac MYHC Genes. We had previously partially sequenced and characterized the β form of the human ventricular MYHC cDNA clone (pHMC3) (15, 24). Isolation of human MYHC genomic clones was first done by screening a cosmid library with a fragment from pHMC3 common to both α- and β-MYHC genes as probe.

A restriction enzyme map of the 51-kb DNA that encompasses the human α- and β-MYHC genes is shown (Fig. 1). The α- and β-MYHC genes are tandemly linked and separated by a 4.5-kb intergenic region, suggesting that these genes may be generated by a gene-duplication event.

Initially, three cosmid genomic clones were isolated, and subsequently, one of these clones, defined as Cos HM-1 in Fig. 1, was identified as the β-MYHC gene. An additional cosmid genomic clone, designated Cos HM-3 (Fig. 1), contained the 3′ region of the α-MYHC gene. The other cosmid clone, designated Cos HM-13 (Fig. 1), which contained the central portion of the α form of the MYHC gene, was isolated by using a 1.9-kb *Hin*dIII and *Xba* I fragment (pE7HX) of Cos HM-1 as probe. The genomic clone that contained the 5′ end of the β-MYHC gene was isolated from the human bacterio-

Abbreviations: MYHC, myosin heavy chain; -MYHC and β-MYHC, α and β forms of MYHC, respectively.

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04535).

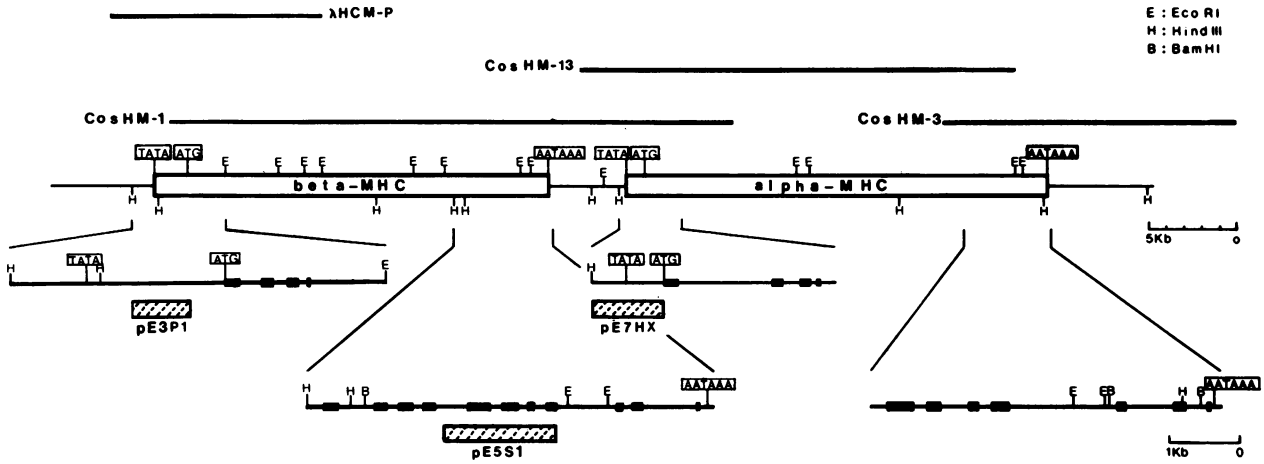

FIG. 1. Restriction endonuclease map of human cardiac MYHC (MHC) genomic clones. Relative positions of the four overlapping clones are given in the 5′ to 3′ orientation of the chromosomal map. Locations of the putative "TATA" box, ATG initiation codon, and AATAAA polyadenylylation signals are indicated. The probes used in the chromosome walks and in the genomic Southern blot hybridizations are shown as dashed boxes. Thick bars represent exons tandomly linked to introns.

phage genomic library using a 0.8-kb *Pst* I fragment (pE3P1) of Cos HM-1 and designated λHCM-P (Fig. 1).

The structure of the cardiac MYHC genes was reaffirmed by Southern blot analyses as shown in Fig. 2. Blot hybridization analysis with a 1.8-kb *Sac* I fragment (pE5S1) derived from the genomic clone Cos HM-1 as probe gave only two positive bands. The *Eco*RI digestion yielded a 13-kb fragment for the α-MYHC gene and a 4.3-kb fragment for the β-MYHC gene; *Bam*HI digestion produced 8 kb for α-MYHC and 13.3 kb for β-MYHC; and *Hin*dIII digestion produced 8 kb for α-MYHC and 7.3 kb for β-MYHC. When the 0.8-kb *Pst* I fragment (pE3P1) and 1.9-kb *Hin*dIII and *Xba* I fragment (pE7HX) of clone Cos HM-1 were used as probes, the *Eco*RI digest yielded 15-kb and 12-kb fragments, *Bam*HI yielded 17-kb and 13.3-kb fragments, and *Hin*dIII yielded 13-kb and 15.5-kb fragments, respectively (data not shown). These findings confirmed that the restriction enzyme map constructed from this cloned DNA described above reflects genomic organization of human cardiac MYHC genes.

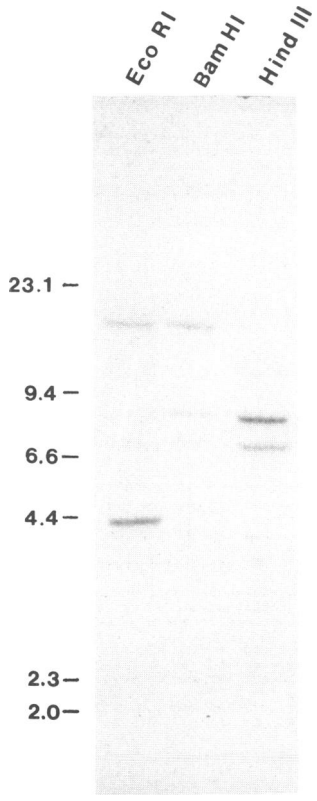

FIG. 2. Southern blot hybridization analysis of human genomic DNA. Ten micrograms of human genomic DNA was digested with *Eco*RI, *Bam*HI, or *Hin*dIII and analyzed by Southern blotting as described. pE5S1 fragment was used as probe. Markers (in kb) were *Hin*dIII fragments of phage.

Sequences of the Cardiac MYHC Genes. To identify the α- and β-MYHC genes, we initially sequenced the isolated DNA fragment that was hybridized to the 3′ region of pHMC3 (15). Subsequent identification of cardiac α-MYHC and β-MYHC genes was done by sequencing the 3′-untranslated region of each clone. Fig. 3 shows the partial nucleotide sequences of human cardiac α- and β-MYHC-encoding genes. The nucleotide sequences of the promoter region and the 15 exons as well as introns of the β-MYHC gene were determined. The nucleotide sequence of the β-MYHC-encoding gene was 100% homologous to the cardiac β-MYHC cDNA clone (pHMC3). Because the β and α forms of the cardiac MYHC genes have been shown to be tandemly linked, we proceeded to sequence the promoter region of α-MYHC-encoding gene in the same cosmid genomic clone, Cos HM-1. The nucleotide sequences of the 11 exons as well as introns of the α-MYHC gene were also determined. The sequences of these regions are very homologous to their counterparts in the rat and hamster MYHC genes (13, 25). Comparison of the nucleotide sequences of the first coding exon between the human α- and β-MYHC genes showed 78% identity. Although the α- and β-MYHC have specific ATPase activity, sequence identity between these two genes is even more pronounced (96%) for the region around the ATP-binding site.

Clone Cos HM-1 has few nucleotide and deduced-amino acid mismatches with that of the human MYHC-encoding gene designated λ-HMHC8 reported by Saez *et al.* (19) in the 5′-flanking region and the first coding exon of α-MYHC. Eighty-eight nucleotides are missing at nucleotide position 756 of λ-HMHC8 (19); the first coding exon consists of 67 amino acids in clone Cos HM-1, as compared with 66 amino acids in λ-HMHC8 (19). Our results are further substantiated by the first 13 amino acids of the α-MYHC-encoding gene, which are identical to that of the rat (13), whereas λ-HMHC8 has three mismatches. The nucleotide sequences reported here have been validated by repeated sequencing of these regions.

DISCUSSION

A total of 18 kb of the 51 kb of the human cardiac MYHC genes have been sequenced; the purpose of sequencing these 18 kb is to elucidate the regulatory regions of the cardiac MYHC genes. Graphic matrix analysis of the 5′-flanking region of the rat and human cardiac α-MYHC genes showed remarkable similarity for 600 nucleotides upstream of the putative TATA box. There is no similarity for the next 1000

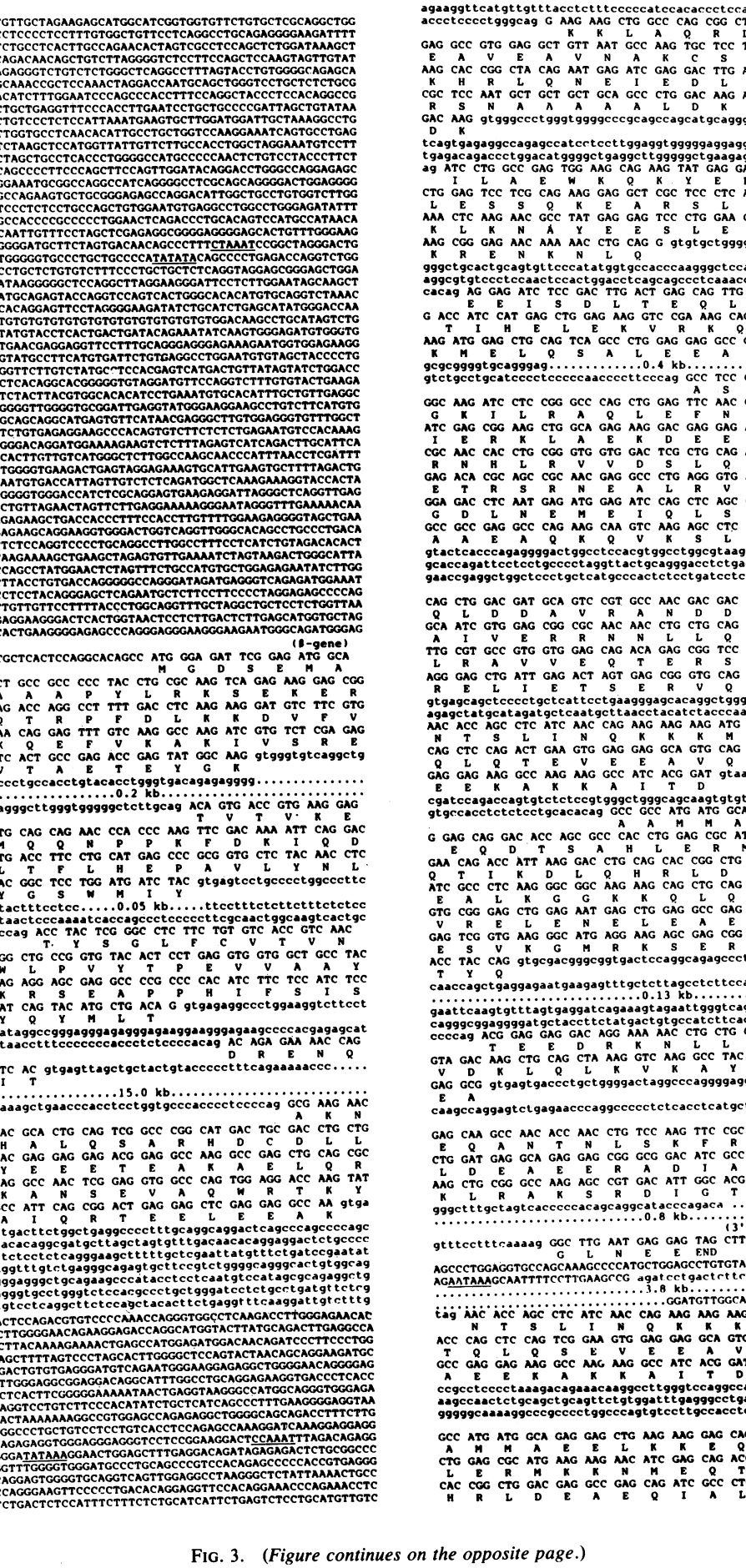

FIG. 3. (*Figure continues on the opposite page.*)

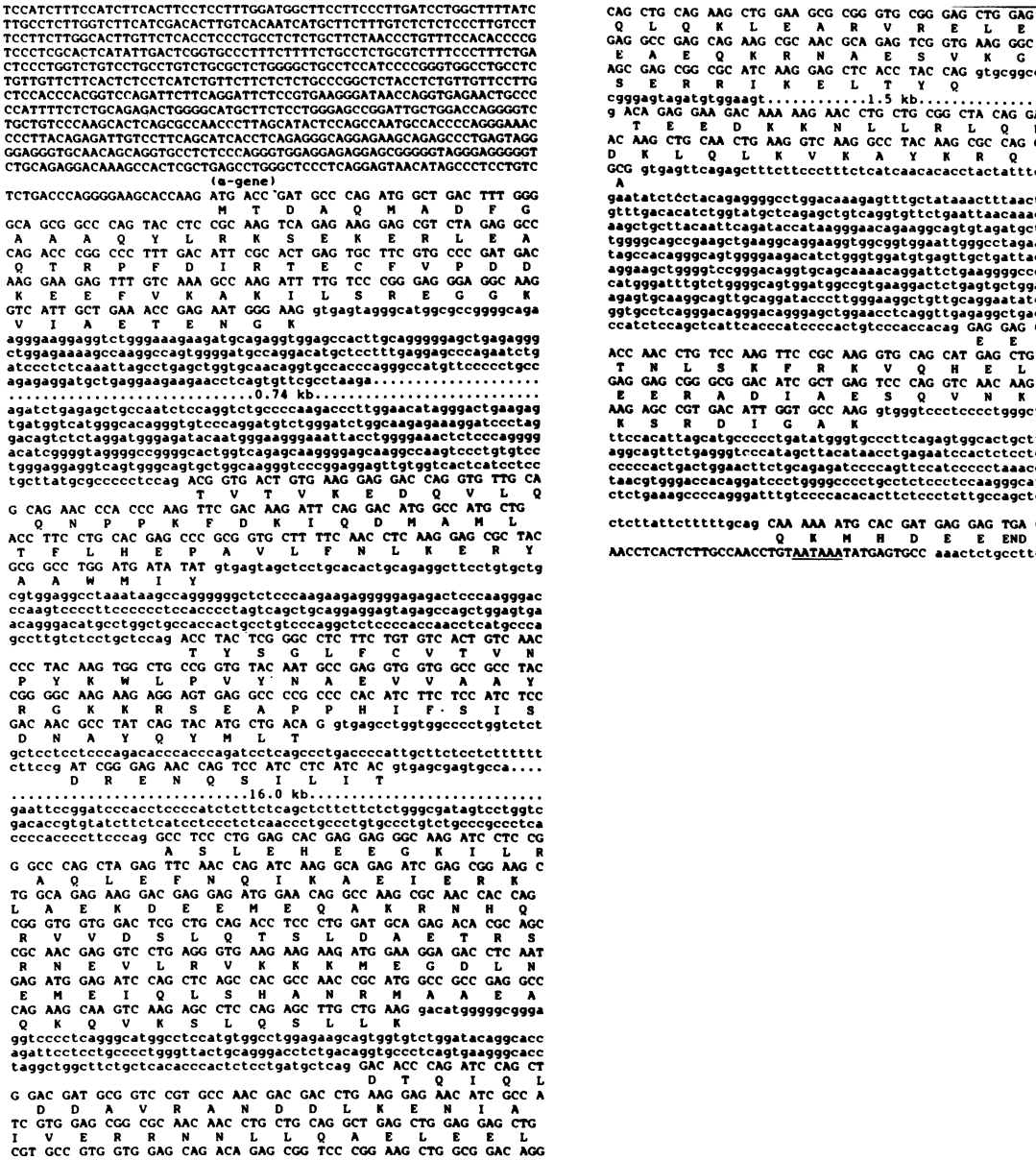

FIG. 3. Partial nucleotide sequence of the human cardiac α- and β-MYHC genes. The nontranscribed strand is displayed in 5′ to 3′ orientation. The exon and flanking sequences are given in uppercase letters. Amino acids encoded within the exons are labeled in the one-letter code beneath the second nucleotide of each codon. CAAT and TATA boxes and AATAAA polyadenylylation signal are underlined. The putative cap site of the mRNA is 29 to 31 nucleotides downstream of the TATA box.

nucleotides toward ATG (data not shown). Conservation of sequences in the 5′-flanking region of these genes may indicate putative regulatory sequences. We discovered, however, that the 5′-flanking region of the human α- and β-MYHC genes diverge extensively from one another. Neither human nor rat cardiac α-MYHC 5′-flanking region showed homology with the human β-MYHC sequence. These results suggest that the human cardiac α- and β-MYHC genes are regulated independently. Comparisons of the 5′-flanking region of β-MYHC gene of other mammals as well as genes under similar control may give more information about the importance of these regions.

The nucleotide sequences in Fig. 3 were used to generate the corresponding amino acid sequences of the α- and β-MYHC genes. These amino acid sequences, as well as the amino acid sequences of rat cardiac (3) and embryonic skeletal (26) and rabbit MYHC proteins (27) are compared and presented (Fig. 4). The first coding exon of the three sequenced cardiac genes encode 67 amino acids. Of these 67, there were only 6 dissimilar amino acids between rat and human α-form MYHC genes, representing 91% identity, whereas there were 13 dissimilar amino acids between the α- and β-form human MYHC genes, representing 81% identity. Within the amino acids encoded by the first coding exon, the magnitude of change between the two cardiac α isoforms (human and rat) is much smaller than between the α and β forms of human cardiac MYHC genes. The sequencing in the first coding region may be characteristic of either α or β isoform, considering the degree of conservation between rat and human sequencing.

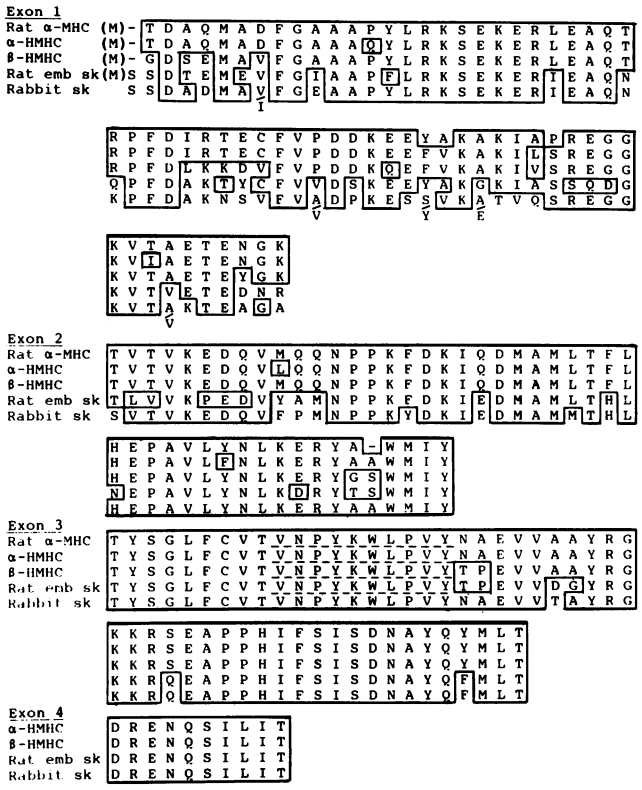

FIG. 4. Comparison of the amino acid sequences from the first four coding exons of the cardiac and skeletal MYHC genes. The deduced amino acid sequences from human α- and β-MYHC-encoding genes are compared with the amino acid sequences of rat cardiac MYHC (3), rat embryonic skeletal MYHC (26), and rabbit skeletal MYHC (27). Gaps represent deletion of an amino acid. Identical residues are boxed. The ATP-binding sites are indicated by dashed lines. One-letter amino acid code is used.

Tong and Elzinga (27) proposed that at least part of the ATP-binding site is located in the third coding exon (Fig. 4), based on the charge distribution of the peptide sequence within this region. α-MYHC, with its high ATPase activity, and β-MYHC, with its low ATPase activity, showed striking sequence conservation in and around this region. Although α- and β-MYHC are needed for optimum ATPase activity, this region of conserved sequence suggests a minimum requirement for an active ATPase domain.

Because there were some differences between the nucleotide sequences of the β forms of the cardiac cDNA clone (pHMC3) (15) and the skeletal cDNA clone (pSMHCZ) (28), we decided to characterize the β-MYHC-encoding gene (Fig. 3). The sequence of the β-MYHC gene reveals that the reported 21-nucleotide divergence for skeletal MYHC occurs in a mid-exon region. It is unlikely that a single exon could be spliced into two different transcripts. Thus, we suggest that only one gene exists for β-MYHC, which is expressed in cardiac muscle and slow-twitch skeletal muscle (20). Southern blot analysis (Fig. 2) confirms our findings by exhibiting only two bands that correspond to the identified cardiac α- and β-MYHC genes. The expression of a single β-MYHC gene in cardiac muscle and slow-twitch skeletal muscle may be due to either the sharing of a common muscle-gene regulatory mechanism or the coexpression of genes that have a complex of distinct cardiac-specific and slow skeletal-specific control elements that modulate muscle-specific gene expression.

We thank Dr. Y. Takihara for technical help to obtain the λHCMP clone from the human genomic library of Dr. T. Maniatis (Harvard University). The typing assistance of Ms. S. Liem and the technical assistance of Ms. A. Madapallimattam are gratefully appreciated. This study was supported by grants from the National Institutes of Health, the Ontario Heart and Stroke Foundation, and the Medical Research Council of Canada.

1. Robbins, J., Freyer, G. A., Chisholm, D. & Gilliam, T. C. (1982). *J. Biol. Chem.* **257,** 549–566.
2. Wydro, R. M., Nguyen, H. T., Gubits, R. M. & Nadal-Ginard, B. (1983) *J. Biol. Chem.* **258,** 670–678.
3. Mahdavi, V., Periasamy, M. & Nadal-Ginard, B. (1982) *Nature (London)* **297,** 659–665.
4. Sinha, A. M., Umeda, P. K., Kavinsky, C. J., Rajamanickam, C., Hsu, H. J., Jakovcic, S. & Rabinowitz, M. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5847–5851.
5. Umeda, P. K., Kavinsky, C. J., Sinha, A. M., Hsu, H. J., Jakovcic, S. & Rabinowitz, M. (1983) *J. Biol. Chem.* **258,** 5206–5214.
6. Lompre, A. M., Nadal-Ginard, B. & Mahdavi, V. (1984) *J. Biol. Chem.* **259,** 6437–6446.
7. Hoh, J. F. Y., McGrath, P. A. & Hale, H. T. (1978) *J. Mol. Cell. Cardiol.* **10,** 1053–1076.
8. Lompre, A. M., Mercadier, J. G., Wisnewsky, C., Bonveret, P., Pantaloni, D., D'Albis, A. & Schwartz, K. (1981) *Dev. Biol.* **84,** 286–291.
9. Everett, A. W., Clark, W. A., Chizzonite, R. A. & Zak, R. (1983) *J. Biol. Chem.* **258,** 2421–2425.
10. Mahdavi, V., Izumo, S. & Nadal-Ginard, B. (1987) *Circ. Res.* **60,** 804–814.
11. Lompre, A. M., Schwartz, K., D'Albis, A., Lacombe, G., Thiem, N. V. & Swynghedaum, B. (1979) *Nature (London)* **282,** 105–107.
12. Gorza, L., Mercadier, J. J., Schwartz, K., Thornell, L. E., Sartore, S. & Schiaffino, S. (1984) *Circ. Res.* **54,** 694–702.
13. Mahdavi, V., Chambers, A. P. & Nadal-Ginard, B. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2626–2630.
14. Friedman, D. J., Umeda, P. K., Sinha, A. M., Hsu, H. J., Jakovcic, S. & Rabinowitz, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3044–3048.
15. Jandreski, M. A. & Liew, C. C. (1987) *Hum. Genet.* **76,** 47–53.
16. Kurabayashi, M., Tsuchimochi, H., Komuro, I., Takaku, F. & Yazaki, Y. (1988) *J. Clin. Invest.* **82,** 524–531.
17. Lichter, P., Umeda, P. K., Levin, J. E. & Vosberg, H. P. (1986) *Eur. J. Biochem.* **160,** 419–426.
18. Catanzaro, D. F. & Morris, B. J. (1986) *Circ. Res.* **59,** 655–662.
19. Saez, L. J., Gianola, K. M., McNally, E. M., Feghali, R., Eddy, R., Shows, T. B. & Leinwand, L. A. (1987) *Nucleic Acids Res.* **15,** 5443–5459.
20. Liew, C. C., Takihara, K. Y., Jandreski, M. A., Liew, J. & Sole, M. J. (1988) in *Sarcomeric and Non-Sarcomeric Muscles: Basic and Applied Research Prospects for the 90s*, ed. Carraro, U. (Unipress Padova, Italy), pp. 11–17.
21. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15,** 687–701.
22. Smith, A. J. H. (1980) *Methods Enzymol.* **65,** 560–580.
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
24. Jandreski, M. A., Sole, M. J. & Liew, C. C. (1987) *Hum. Genet.* **77,** 127–131.
25. Liew, C. C. & Jandreski, M. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3175–3179.
26. Strehler, E. E., Strehler-Page, M. A., Perriard, J. C., Periasamy, M. & Nadal-Ginard, B. (1986) *J. Mol. Biol.* **190,** 291–317.
27. Tong, S. W. & Elzinga, M. (1983) *J. Biol. Chem.* **258,** 13100–13110.
28. Saez. L. J. & Leinwand, L. A. (1986) *Nucleic Acids Res.* **14,** 2951–2969.