

Characterization of a mammalian smooth muscle myosin heavy-chain gene: Complete nucleotide and protein coding sequence and analysis of the 5' end of the gene

PHILIP BABU*, CAROL KELLY, AND MUTHU PERIASAMY†

Department of Physiology and Biophysics, University of Vermont College of Medicine, Given Building, Burlington, VT 05405

Communicated by Andrew Huxley, August 26, 1991

ABSTRACT The purpose of this study was to characterize the complete cDNA sequence encoding the rabbit smooth muscle myosin heavy chain (MHC) and determine the exon/intron organization at the 5' end of the corresponding gene. The full-length cDNA sequence of 6644 base pairs encoding a protein of 1972 amino acids was generated from two cDNA clones: PBRUC1 (≈6.3 kilobases), isolated from a rabbit uterus cDNA library, and PBRU-PCR33 (420 base pairs), produced by primer extension and PCR amplification. Compared with the chicken smooth muscle MHC sequence [Yanagisawa, M., Hamada, Y., Katsuragawa, Y., Imamura, M., Mikawa, T. & Masaki, T. (1987) *J. Mol. Biol.* 198, 143–157] the rabbit MHC shares about 90% amino acid identity in the S1 globular head region but shows a striking sequence divergence at the junction between the 25-kDa and 50-kDa proteolytic fragments of the functionally important S1 head domain. Genomic cloning shows that the rabbit smooth muscle MHC gene is large and has an unusual exon/intron organization at the 5' end. The first eight contiguous exons are located within a region of at least 70 kilobases of genomic DNA. Some introns span several kilobases of DNA and others at the 5' end show a high degree of intron conservation in the Mg²⁺-ATPase domain when compared with more distantly related sarcomeric MHC genes. Primer extension and S1 nuclease mapping analysis demonstrate that transcription initiates from a single site in the rabbit smooth muscle MHC gene.

The contractile properties of smooth muscle cells show distinct differences compared to striated muscle. In smooth muscle cells, contraction is regulated by a Ca²⁺/calmodulin-dependent myosin light-chain kinase. The tension produced during contraction can be equal to or greater than that produced in striated muscle despite the presence of a smaller myosin/actin ratio in smooth muscle cells. Furthermore, contractions in smooth muscle cells can be more economical since tension can be maintained in the absence of ATP splitting. To determine how myosin heavy chain (MHC) contributes to the unique contractile properties of smooth muscle cells, we began a detailed analysis of the structure and expression of the smooth muscle MHC gene.

Previous work on smooth muscle MHC expression showed that two polypeptides with estimated molecular weights of 204,000 (SM1) and 200,000 (SM2) could be identified by SDS/polyacrylamide gel electrophoresis in a number of smooth muscle tissues (1). However, cDNA cloning of myosin from chicken gizzard (2) and rabbit uterus (3) suggested that MHC was encoded by a single gene in smooth muscle cells. To determine the structure and complexity of MHC isoforms in smooth muscle cells, we isolated two distinct cDNA clones and demonstrated that they code for SM1 and SM2 MHC isoforms (4, 5). Subsequently we showed that

SM1 and SM2 isoforms were alternatively spliced products of the same gene (5). It is unknown whether specialized structural or functional domains in the MHC molecule or differences in SM1 and SM2 isoform expression contribute to the unique contractile properties of smooth muscle cells.

In the present study we characterized the complete amino acid and nucleotide structure of the rabbit smooth muscle MHC (SM1) and determined the 5'-end exon/intron organization spanning putative functional domains of the MHC gene.‡ Our studies show that the rabbit smooth muscle MHC gene is large and has an unusual exon/intron organization at the 5' end. Some introns span several kilobases of genomic DNA and others show a high degree of intron conservation in the Mg²⁺-ATPase domain when compared with more distantly related sarcomeric MHC genes. The mammalian (rabbit) and avian (chicken) smooth muscle MHC share about 90% amino acid homology in the S1 globular head region but demonstrate a high sequence divergence at the junction between the 25-kDa and 50-kDa proteolytic fragments of the functionally important S1 head domain.

METHODS

Construction and Screening of Rabbit Uterus Smooth Muscle Library. The construction of a full-length rabbit uterus cDNA library [size range, 5–10 kilobases (kb)] in λZAP has been described (6). The library was screened with a ³²P-labeled DNA fragment from the 5' end of the rat aorta cDNA clone RAMHC 21 (5). Restriction fragments from the longest cDNA clone, PBRUC1, ≈6.3 kb, were subcloned into pUC18/19 and M13mp18/19 vectors and sequenced completely by the dideoxy method using Sequenase (United States Biochemical) (5).

PCR Cloning of the 5' End of the Rabbit Smooth Muscle MHC mRNA. To obtain the missing 5'-end cDNA sequence that was not present in PBRUC1, the technique for rapid amplification of cDNA ends (RACE; ref. 7) was used in conjunction with the polymerase chain reaction (PCR). Poly(A)⁺ mRNA (8) from adult rabbit uterus was reverse-transcribed with 200 units of Superscript (RNase H⁻ Moloney murine leukemia virus reverse transcriptase; Bethesda Research Laboratories) using a ³²P-labeled 26-mer primer (5'-TGCACAACCTGAGGGAGAGGTTCTC-3') from the 5' end of PBRUC1. The extended product mixture was treated with 10 units of RNase H (Boehringer Mannheim), resuspended in alkaline loading buffer, and purified following denaturing 6% polyacrylamide gel electrophoresis. The sin-

Abbreviations: MHC, myosin heavy chain; RACE, rapid amplification of cDNA ends.

*Present address: Department of Physiology, University College London, Gower Street, London WC1E 6BT, United Kingdom.

†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M77812).

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.

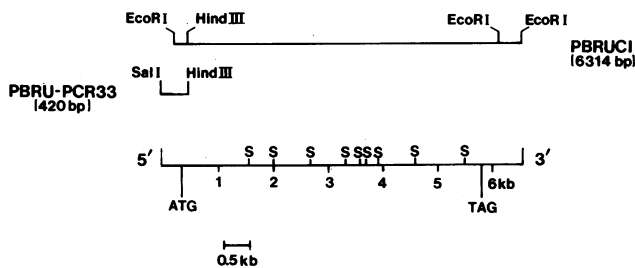


Fig. 2. Restriction map (S, *Sac* I) of complete cDNA sequence for rabbit uterus smooth muscle MHC. Primer extension and PCR cloning were used to isolate the 5' untranslated sequence represented by cDNA clone PBRU-PCR33. The amplified DNA (420 bp) includes 31 bp of primer sequence at the 5' end, 96 bp of 5' untranslated sequence, 234 bp of coding sequence, and 59 bp of 3' sequence that overlaps with cDNA clone PBRUC1.

using 35 cycles of amplification (Perkin-Elmer). DNA sequence analysis of this MHC cDNA clone (PBRU-PCR33) revealed that it contained 5' untranslated sequence plus the missing coding sequence.

Isolation of Genomic Clones Corresponding to the 5' End of the Rabbit Smooth Muscle MHC Gene. A rabbit genomic library constructed in EMBL3 (Clontech) was screened both with the PBRU-PCR33 cDNA subclone and with restriction fragments from the 5' end of PBRUC1. Several positive clones were identified and characterized by Southern blotting. Selected genomic subclones that hybridized to the cDNAs were sequenced in both orientations to map exon/intron boundaries at the 5' end of the MHC gene.

Identification of Transcription Initiation Site. A combination of primer extension and S1 nuclease mapping was used to locate the transcription initiation site in the MHC gene. Two oligonucleotides, A (18-mer) and B (19-mer), were synthesized to complementary sequences in the 5' untranslated region of the smooth muscle MHC mRNA (Fig. 1). The primers were 5' end-labeled with 32 P and annealed to RNA (see Fig. 5) in 40 mM Pipes, pH 6.8/1.25 mM EDTA, pH 8.0/125 mM NaCl/75% (vol/vol) formamide for 60 min at 42°C. Hybrids were ethanol-precipitated and reverse-transcribed with 200 units of Superscript in a mixture containing 0.06 μ g of actinomycin D (9). Extension products were analyzed in denaturing 8% or 20% polyacrylamide gels.

For S1 nuclease mapping, a 900-bp *Pst* I fragment from genomic clone ARG-4 (see Fig. 4) containing the first exon was subcloned into M13mp18 and uniformly labeled as described (10). The labeled DNA was cut with *Sac* I and *Pst* I to release a 220-bp fragment that was purified in a strand-separating 5% polyacrylamide gel. The single-stranded probe was hybridized (4) to rabbit uterus poly(A)⁺ RNA and to total RNA samples from different smooth muscle tissues. After S1 nuclease treatment, protected fragments were phenol-

extracted and analyzed in denaturing 8% polyacrylamide gels.

RESULTS AND DISCUSSION

Characterization of Full-Length cDNA Sequence for Rabbit Smooth Muscle MHC. A cDNA library made from adult rabbit uterus was screened with a 32 P-labeled DNA fragment from the 5' end of the rat aorta cDNA clone RAMHC21 (5). Several positive MHC cDNA clones were identified and the longest cDNA (PBRUC1, \approx 6.3 kb) was completely sequenced (Fig. 1). The missing sequence from the 5' end (encoding \approx 80 amino acids) was obtained using primer extension in conjunction with PCR in the RACE protocol. The complete nucleotide sequence of rabbit smooth muscle MHC determined from the two overlapping clones PBRUC1 (\approx 6.3 kb) and PBRU-PCR33 (420 bp) is shown in Fig. 1. The full-length rabbit uterus MHC cDNA is 6644 bp long and encodes a protein of 1972 amino acids, which is similar in size to the 1979-amino acid sequence reported for chicken gizzard MHC (2). The cDNA clone PBRUC1 represents a longer version of SMHC 40 reported previously (4) and corresponds to the SM1-type (204-kDa) MHC. The PCR product (PBRU-PCR33) encoding the 5' untranslated and missing coding sequences is shown in Fig. 2. The 3' end of PBRU-PCR33 contains 59 nucleotides that are identical to and overlap with the 5' end of cDNA clone PBRUC1. The missing coding sequence and start codon were identified by comparison with chicken gizzard (2) and rat skeletal embryonic (11) MHC genes. Based on this comparison, PBRU-PCR33 encoded 78 amino acids including the start codon and the complete 5' untranslated sequence of 96 nucleotides. This was verified by primer extension and S1 nuclease mapping analyses (see below).

Rabbit Uterus and Chicken Gizzard MHC Deduced Protein Sequences Diverge at the 25-kDa/50-kDa Junction in the Globular S1 Head Region. Fig. 3 compares the amino acid sequence corresponding to part of the S1 head region of the rabbit uterus MHC with chicken gizzard (2) and chicken nonmuscle (12) MHCs. Overall the amino acid sequences are highly conserved, showing \approx 90% identity. The putative Mg^{2+} -ATPase and actin-binding domains are virtually identical for all three MHCs. However, an interesting difference in amino acid sequence between rabbit and chicken smooth muscle MHCs was evident in the S1 head precisely at the junction of the 25-kDa and 50-kDa proteolytic fragments. Comparison of the two sequences in this region shows that the two MHCs are homologous up to amino acid 205 but then diverge completely from amino acid 206 with respect to rabbit uterus MHC. From this point of divergence, the rabbit uterus MHC encodes 6 unique amino acids whereas chicken gizzard MHC encodes a stretch of 13 unique amino acids before the high degree of homology is restored (Fig. 3). The extra amino

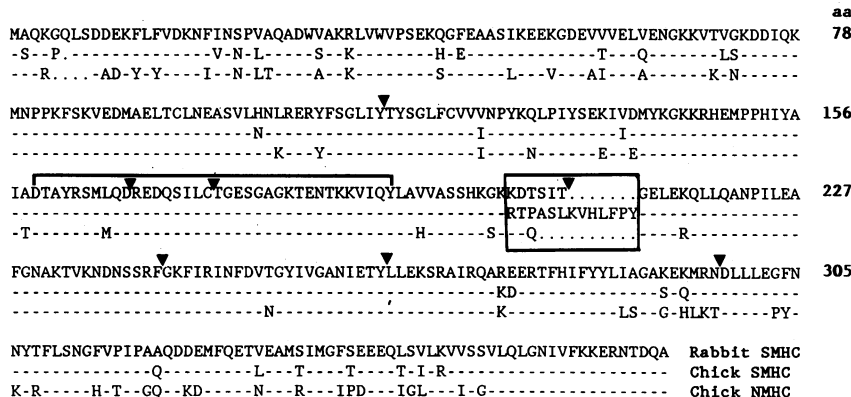


Fig. 3. Comparison of deduced amino acid (aa) sequences of S1 globular head region of rabbit uterus smooth muscle MHC (SMHC), chicken gizzard SMHC (2), and chicken nonmuscle MHC (NMHC) (12). Dashes represent identical sequence, dots represent spaces inserted for optimal alignment of amino acid residues. Boxed area identifies unique region in MHC molecule where sequences diverge. Bracket shows putative Mg^{2+} -ATPase domain (residues 159-194) with respect to rabbit uterus sequence. Solid triangles represent exon/intron boundaries identified in the rabbit smooth muscle MHC gene (see Fig. 4).

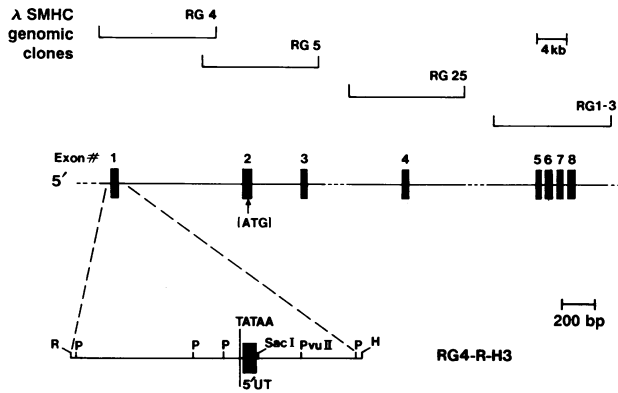


FIG. 4. Exon/intron organization at the 5' end of the rabbit smooth muscle MHC (SMHC) gene. Black boxes represent exons and horizontal line represents intervening sequence; dotted horizontal line represents regions where no overlapping genomic clones have been identified and therefore represents intervening sequence of unknown length. A detailed restriction map of the nucleotide sequences immediately upstream and downstream of the first noncoding exon is shown. R, *EcoRI*; P, *Pst I*; H, *HindIII*; 5' UT, 5' untranslated.

acid sequence found in chicken gizzard MHC is not present in chicken nonmuscle MHC.

In the MHC molecule, the globular S1 head region is known to contain ATPase activity and actin-binding properties (13). Tryptic digestion of the S1 head results in the production of three proteolytic fragments of about 25, 50, and 23 kDa. In general the mammalian MHCs (sarcomeric and nonsarcomeric) show high amino acid homology in the S1

region, so the finding of divergent sequences (Fig. 3) between the two smooth muscle MHCs at the 25-kDa/50-kDa junction may be structurally and functionally significant. Recently it was proposed (14) on the basis of S1 nuclease mapping experiments that the chicken smooth muscle MHC gene was alternatively spliced at the 5' end to generate vascular and intestinal isoforms. The putative splicing region was localized to the 25-kDa/50-kDa boundary in the S1 head, thus implicating heterogeneity in the ATPase domain as a factor involved in the functional differences between vascular and visceral smooth muscle cell types. Furthermore, a remarkable arrangement of duplicated exons is known to encode *Drosophila* muscle MHC (15), and alternative splicing of these exons can result in multiple potential MHC isoforms, some of which are known to be expressed in a stage- and tissue-specific manner. The results shown in Fig. 3 provide direct evidence that the smooth muscle MHC 25-kDa/50-kDa junction differs between species and may represent a functionally important domain in the MHC molecule.

Exon/Intron Organization of the 5' End of the Rabbit Smooth Muscle MHC Gene. Two overlapping genomic clones, λ RG-4 and λ RG-5, hybridized with PCR-amplified cDNA clone PBRU-PCR33, indicating that these genomic clones contained the 5' end of the MHC gene (Fig. 4). Mapping of the exon/intron boundaries in this region of the gene was based on the consensus GT . . . AG for splice junctions and comparison of the genomic sequence with cDNA sequence. DNA sequence analyses revealed an unusual exon/intron organization at the 5' end (Fig. 4) in that the first eight exons (equivalent to only 1.0 kb of the 6.6-kb mRNA) are located within a region of at least 70 kb of genomic DNA. The precise intronic distance between exons

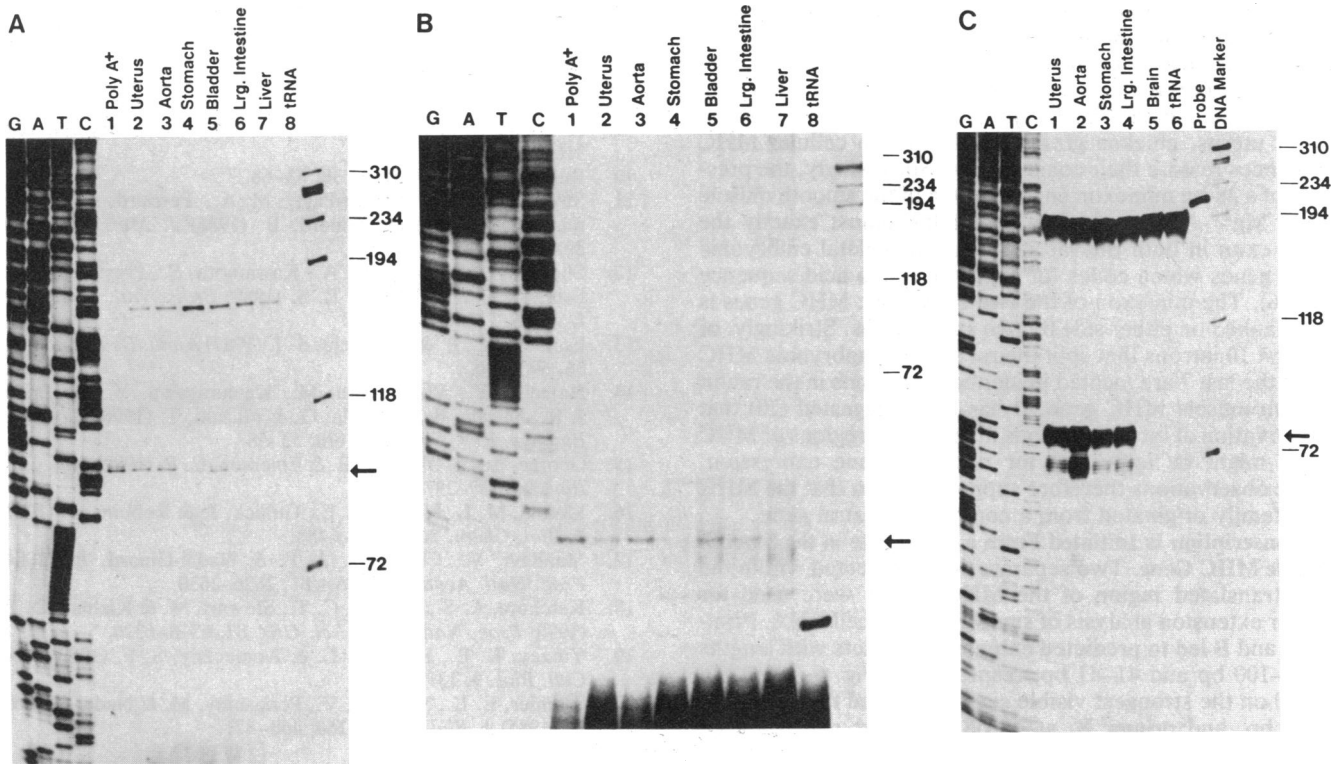


FIG. 5. Identification of transcription initiation site for rabbit smooth muscle MHC gene. The source of RNA for primer extension and S1 nuclease mapping was poly(A)⁺ RNA (1 μ g) from rabbit uterus and total RNA (30 μ g) from adult rabbit tissues. (a) Primer extension with primer A produces a fragment with a predicted size of 98–100 bp in denaturing 8% polyacrylamide gel. A band is evident at \approx 180 bp but is nonspecific since it is not present in poly(A)⁺ RNA or tRNA samples. No extension product is seen for liver RNA or tRNA, which act as negative controls. Sequencing ladder (lanes G, A, T, and C) is from 5' end of PCR33. (b) Primer extension with primer B produces a fragment with a predicted size of 41–43 bp in a denaturing 20% polyacrylamide gel. (c) S1 nuclease mapping of 5' end of rabbit smooth muscle MHC gene. The probe protects a major fragment with a predicted size of 78–84 bp. No protection is evident in brain and tRNA samples, which act as negative controls. A nonspecific band is present at \approx 200 bp in all RNA samples and probably results from incomplete strand separation of the DNA probe.

3 and 4 and between exons 4 and 5 is unknown, since two rounds of genomic library screening and Southern blotting failed to identify overlapping genomic clones in these regions. The unusually large sizes of the intervening sequences at the 5' end of the smooth muscle MHC gene are not evident in other characterized vertebrate sarcomeric MHC genes (11, 16, 17).

Primer extension cloning (7) of the 5' end of the smooth muscle MHC cDNA led to the finding that the 5' untranslated sequence (95 bp) in the MHC gene was encoded in two exons separated by >20 kb of intervening sequence (Fig. 4). The first noncoding exon is 78 bp long, and the remaining 17 bp of 5' untranslated sequence is contained in the first coding exon (exon 2, 366 bp). Other MHC genes also show this type of split organization in the 5' untranslated region (11, 15–17). In the case of *Drosophila* nonmuscle MHC (18), one of two possible 5' untranslated sequences participates in 5'-end alternative splicing to generate nonmuscle MHC isoform diversity. Although the significance of large intronic sequences of the magnitude shown in Fig. 4 are unknown, it has been shown that regulatory sequences do reside in the 5' introns of some muscle genes (19).

The putative Mg^{2+} -ATPase domain is encoded in exons 3–5; exon 4 is a 28-bp minixon that encodes 8 central amino acids of the putative Mg^{2+} -ATPase domain. The beginning and end of the minixon are formed by two split codons. The second half of the Mg^{2+} -ATPase domain is encoded in exon 5, whose coding sequence ends at the 25-kDa/50-kDa junction in the S1 head region. This is the region of the MHC molecule that shows divergence from the chicken gizzard sequence (2). Our results show that the exon/intron boundaries are conserved between rabbit smooth muscle MHC (exons 3–6) and rat sarcomeric MHC (exons 5–7) over the Mg^{2+} -ATPase and 25-kDa/50-kDa domains. In this region the smooth muscle MHC and the sarcomeric MHC showed 65% amino acid sequence identity, thus emphasizing the importance of this conserved region for myosin function. The beginning of exon 6 (Fig. 4) represents the point at which the rabbit uterus, chicken gizzard, and chicken cellular MHC sequences restore their conservation. Surprisingly, the presence of a 28-bp minixon encoding part of the smooth muscle MHC Mg^{2+} -ATPase domain resembles almost exactly the 28-bp exon in both the rat and chicken skeletal embryonic MHC genes which codes for a similar amino acid sequence (11, 16). The minixon of the two embryonic MHC genes is also flanked on either side by two split codons. Strikingly, of the first 10 introns that split the rat skeletal embryonic MHC gene, the last 7 are located in identical positions in the rabbit smooth muscle MHC gene. It has been suggested (20) that conservation of intron positions in the 5'-end regions of MHC genes might be important for regulated gene expression. These observations therefore support the idea that the MHC gene family originated from a common ancestral gene.

Transcription Is Initiated From a Single Site in the Smooth Muscle MHC Gene. Two separate primers located within the 5' untranslated region of the MHC cDNA were used for primer extension analysis of smooth muscle cell RNA. Primers A and B led to predicted extension products with lengths of 98–100 bp and 41–43 bp, respectively (Fig. 5 A and B). Based on the strongest visible signal observed for primer A, at 98 bp, and primer B, at 41 bp, the start site could be

assigned to the guanine residue that appears as nucleotide +2 in the composite cDNA sequence (Fig. 1). It is possible that the adenine nucleotide at +1 in the cDNA sequence is a cloning artifact from the PCR reaction, although an adenine nucleotide does occur in the corresponding genomic sequence (data not shown). The next upstream base in the genomic sequence is also a guanine, indicating that any member of the GAG triplet could be the true start site. S1 nuclease mapping (Fig. 5C) with a genomic fragment containing the 5' untranslated region produced a partially protected fragment of 78–84 bp that corresponded to the same position identified by primer extension analysis. Further, Fig. 5 indicates that in all smooth muscle samples studied, transcription is initiated at a single site in the smooth muscle MHC gene. The presence of a canonical TATAAA sequence 26 nucleotides upstream of the putative start site (data not shown) provides indirect evidence that the start site is genuine for the smooth muscle MHC gene.

We thank Dr. Jonathan Lytton for constructing the rabbit uterus cDNA library. This work was supported by National Institutes of Health Grant R29-38355. M.P. is an Established Investigator of the American Heart Association. P.B. was supported by a postdoctoral fellowship from the American Heart Association, Vermont Affiliate.

- Rovner, A. S., Thompson, M. M. & Murphy, R. A. (1986) *Am. J. Physiol.* **19**, C861–C870.
- Yanagisawa, M., Hamada, Y., Katsuragawa, Y., Imamura, M., Mikawa, T. & Masaki, T. (1987) *J. Mol. Biol.* **198**, 143–157.
- Nagai, R., Larson, D. M. & Periasamy, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1047–1051.
- Nagai, R., Kuro-o, M., Babij, P. & Periasamy, M. (1989) *J. Biol. Chem.* **264**, 9734–9737.
- Babij, P. & Periasamy, M. (1989) *J. Mol. Biol.* **210**, 673–679.
- Lytton, J., Zarain Herzberg, A., Periasamy, M. & MacLennan, D. H. (1989) *J. Biol. Chem.* **264**, 7059–7065.
- Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
- Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
- Krug, M. S. & Berger, S. L. (1987) *Methods Enzymol.* **152**, 316–325.
- Burke, J. F. (1984) *Gene* **30**, 63–68.
- Strehler, E. E., Strehler-Page, M. A., Perriard, J. C., Periasamy, M. & Nadal-Ginard, B. (1986) *J. Mol. Biol.* **190**, 291–317.
- Shohet, R. V., Conti, M. A., Kawamoto, S., Preston, Y. A., Brill, D. A. & Adelstein, R. S. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7726–7730.
- Emerson, C. P. & Bernstein, S. I. (1987) *Annu. Rev. Biochem.* **56**, 695–726.
- Hamada, Y., Yanagisawa, M., Katsuragawa, Y., Coleman, J. R., Nagata, S., Matsuda, G. & Masaki, T. (1990) *Biochem. Biophys. Res. Commun.* **170**, 53–58.
- George, E. L., Ober, M. B. & Emerson, C. P. (1989) *Mol. Cell. Biol.* **9**, 2957–2974.
- Molina, M. I., Kropp, K. E., Gullick, J. & Robbins, J. (1987) *J. Biol. Chem.* **262**, 6478–6488.
- Mahdavi, V., Chambers, A. P. & Nadal-Ginard, B. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2626–2630.
- Ketchum, A. S., Stewart, C. T., Stewart, M. & Kiehart, D. P. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6316–6320.
- Yutzey, K. E., Kline, R. L. & Konieczny, S. F. (1989) *Mol. Cell. Biol.* **9**, 1397–1405.
- Strehler, E. E., Mahdavi, V., Periasamy, M. & Nadal-Ginard, B. (1985) *J. Biol. Chem.* **260**, 468–471.