Structural and Transcriptional Analysis of a Chicken Myosin Heavy Chain Gene Subset*

(Received for publication, January 20, 1987, and in revised form, August 17, 1987)

Keith E. Kropp, James Gulick, and Jeffrey Robbins§
From the Department of Pharmacology and Cell Biophysics, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0575

Recently we have isolated a large number of chicken myosin or myosin-like heavy chain genes. Seven of these genes were placed into a subset based upon their hybridization patterns. In the present study, the sequences of the 5' end of one of these genes, myosin heavy chain (MHC) genes, N127, was determined and compared with the 5' end sequences of the other six MHC genes in the subset. The comparison revealed that the three exons encoding the amino terminus of the protein are highly conserved. The sequence analysis shows that a localized correction event occurred in and around a domain of the nucleotide-binding site, as the exon encoding this site and the preceding intron are very highly conserved among the seven genes.

The sequence of the promoter and 5'-untranslated region of N127 is presented. The analogous regions for N124 and N125 have now been sequenced and are also presented. As is the case for all the other known MHC genes, the 5'-untranslated regions are split by large introns. The promoter and 5'-untranslated regions are compared with two previously characterized chicken MHC genes (N116 and N118) to determine the sequence similarities and differences that might underlie the differential expression of the family's members.

To confirm and extend previously published results of the expression of these genes, transcript-specific probes generated from the 5' region of six of the seven genes were used to determine in which muscle(s) the corresponding mRNAs were present. The data show that despite the very close structural homologies, each of the genes for which a unique probe could be prepared exhibits a unique pattern of expression.

The myosin heavy chain (MHC)1 isoforms of vertebrates are encoded by multigene families (1-4) consisting of at least seven genes (1) and possibly more (5, 6). Different members of the family are expressed in specific tissues and at specific times during development (7-11); thus, the tissue- and developmental stage-specific expression of the MHC genes offers an excellent system for studying gene regulation.

The mechanisms involved in the expression of the MHC genes remain largely obscure. Available data indicate that the expression of a particular MHC isoform can be regulated by various hormonal (9, 11-13) and physiological (14) signals and/or by different innervational (15-17) patterns. Although it is likely that gene expression in this family is regulated at both the transcriptional and post-transcriptional levels (18-23), activation of the transcriptional apparatus seems to play the primary role in the expression of the genes (22).

It is known that basal promoter elements for RNA polymerase II are found within the 5'-flanking region of a gene and are involved in transcriptional regulation (reviewed in Ref. 24). One element, often referred to as the Goldberg-Hogness or "TATA" box, is located approximately 31 bases upstream from the start of transcription (25). This element is responsible for the correct transcriptional initiation at the cap site (24, 26-28). Further upstream, between -50 and -105, another basal promoter element is found which may have either a "CAAT" consensus sequence (29, 30) or a GC-rich (GGGCGG or CCGCCC) sequence (31). This element has been shown, by deletion and linker scanning mutations, to play a role in controlling the frequency of initiation (30-36). The GC-rich region can also bind proteins that have the ability to up-regulate gene expression (reviewed in Ref. 37). A third, cis-acting element affecting transcription, the enhancer, may be located either upstream or downstream from the cap site and seems to exert its effect (relatively) independently of distance and/or orientation from the cap site (38, 39). Enhancers have been shown to be tissue-specific (40, 41) and to mediate steroid hormone activation (42).

At a minimum, in order to understand the developmental expression of the MHC genes, it is necessary to know the overall sequence organization of the gene being expressed and be able to detect that gene's specific transcript. Also, it is particularly important to determine the sequences of different promoters, so that the similarities and differences among these sequences can be catalogued and hopefully correlated with the genes' expression. We have addressed these problems by focusing our attention on a group of seven chicken MHC genes (5, 6, 43, 44), which were placed into a subset based upon their structural homologies. Here, we present data characterizing the 5' end of a member of this subset and compare its transcriptional pattern to other subset members. In addition, we compare the promoter and 5'-untranslated regions of five of these genes in an initial attempt to locate structural regions that may be involved in the tissue- and developmental stage-specific expression of the MHC genes.
**EXPERIMENTAL PROCEDURES**

The procedures used in isolating and mapping the genomic clone have been described previously (5).

**Isolation of DNA Templates for Sequencing by the Rapid Deletion Method**—The procedure developed by Dale et al. (45) for generating sequential overlapping clones for DNA sequencing was employed. 80 ng of a 20-base primer, which is complementary to a region of the M13 phage that encompasses the EcoRI site of the cloning region as well as the 10 bases immediately upstream of the EcoRI site, was annealed to 4 µg of single-stranded template DNA. The insert sizes ranged from 0.3 to 4.5 kilobase pairs. After annealing, the DNA was digested with 10 units of EcoRI at 45 °C for 1 h. The digestion products were loaded onto a 0.7% agarose gel and electrophoresed for 2 h to confirm that at least 95% of the DNA had been cleaved by the enzyme. The DNA was then digested in two separate reactions with either 1.5 or 4 units of T4 DNA polymerase enzyme. The DNA was then digested with 10 units of EcoRI at 45 °C for 1 h. The digestion was carried out at 37 °C. Aliquots from each reaction were taken at various time points (2-70 min) and placed at 65 °C for 10 min to inactivate the enzyme. The aliquots were combined with a 7-20-base deoxyriboamidase "tail" added using a unit of terminal deoxynucleotidyl transferase and 3 µl of 10 mM dGTP. The 20-base primer, which has a small poly(dC) tail, was annealed to the DNA. The DNA was ligated using 1 unit of T4 DNA ligase at room temperature for at least 1 h. The reaction was terminated by the addition of 1 µl of 0.5 M EDTA and the volume increased to 30 µl with water. Finally, 1 µl of a 1/10 dilution of the ligation mixture was used to transform 0.3 ml of JM101 competent cells.

The single-stranded DNA templates were isolated using the procedure provided by Amersham Corp. (46) with the following modifications. The template DNA was precipitated by adding 0.5 volume of 7.5 M NH4OAc (pH 5.4) and 2 volumes of absolute ethanol. The DNA was collected by centrifugation at 4 °C for 15 min or at -20 °C overnight. The DNA was collected by centrifugation at 4 °C for 10 min. The pellet was washed with 80% ethanol, dried, and resuspended in 50 µl of water.

The approximate sizes of the resultant deletions were determined by separating the DNA templates on a 0.6% agarose gel for 3-4 h at 100 mA in Tris borate buffer (45). Bacteriophage with and without the insert were used as markers. Template DNAs displaying sequential deletions of approximately 150 bases were sequenced using the dideoxy sequencing method (47).

All sequences were analyzed on an IBM XT using the programs developed by Queen and Korn (48).

**RNA Isolation, S1 Nuclease, and Primer Extensions Analyses**—These procedures were performed as described previously (5, 43).

**Rot Blots**—3 µg of total RNA from each of the muscle sources was dissolved in 30 µl of 10 mM Tris (pH 7.5), 50 mM EDTA, 30 µl of 37% formaldehyde (Fisher) and 20 µl of 20 × SSC (1 × SSC = 0.15 M sodium chloride, 0.015 M sodium citrate) were added and the RNA heated at 60 °C for 15 min. After 700 µl of ice-cold 20 × SSC was immediately added, 600 µl of this was applied to nitrocellulose which had been pre-dried in 20 × SSC; the remaining 400 µl was used to make serial dilutions using 20 × SSC as diluent.

The paper was dried at 80 °C for 1 h. The paper was rinsed in water and dried in the 37 °C oven. The paper was then prehybridized in 5 × SSC, 0.1% sodium dodecyl sulfate for 30 min, after which the [3P]-labeled oligonucleotide was added to a final concentration of 1 × 106 cpm/ml. Hybridization and autoradiography were carried out as described previously (43).

**RESULTS**

**Characterization and Sequencing Strategy**—The EcoRI restriction map of the bacteriophage encoding N127 is shown in Fig. 1. N127 encodes the 5' end of a MHC gene belonging to the subset of seven clones characterized previously (6). The EcoRI fragments of N127 that contain the sequences encoding the amino-terminal end of the protein were determined by hybridization using the 423-base pair PstI/Sau3A fragment of N116, which was shown previously to encode the 5' end of an adult MHC gene (43). The EcoRI fragments were subcloned into pBR325 and more detailed restriction maps obtained. The desired fragments were subcloned into M13mp19 and sequenced. If the subcloned fragment could not be completely sequenced in one reaction and an overlapping fragment could not be found to complete the sequence, then a series of nested deletions was made from the subcloned fragment using the rapid deletion method (see "Experimental Procedures") and the sequences of the deletion clones determined. The sequencing strategy is presented in Fig. 1.

**Comparison of the Nucleotide and Amino Acid Sequences**—The nucleotide and derived amino acid sequences of N127 are compared with other chicken MHC sequences (N116, N118, N101, N122, N124, and N125 (6, 4, 44, 44)) in Fig. 2, A and B. The sequence of clone N122, which is an overlap clone of N29 (5, 6), was used in these comparisons. The sequences have been aligned to maximize the apparent homologies. N116 was chosen arbitrarily as the standard for comparison. The sequences can be placed into two distinct groups based upon their homologies in the coding or translated region; N116, N101, and N122 are in one group, and N118, N125, N124, and N127 are in the other. These two groups can be subdivided further on the basis of their apparent homologies. For example, the sequences of the N101/N122 subgroup, as well as the N124/N127 subgroup, are highly conserved, whereas the sequences of the N118/N125 subgroup are almost completely conserved. Furthermore, the C repeats occur among the four subgroups and not within a particular subgroup. For example, when the first coding exons (exon 3) of N118 and N125 are compared, starting from the translational initiation codon, only four nucleotide differences and three amino acid substitutions are found (if the first four codons are omitted). The reasons for these subgroups are unknown but are consistent with the hypothesis that this gene family arose through serial gene duplication events.

**The Exons**—The first two exons shown in Fig. 2 encode part of the 5'-untranslated region of the gene and have diverged to a much greater extent than the sequences which encode the translated portions of the mRNAs. The reason for the lack of conservation is unknown but may reflect the differential expression of the genes in terms of specific sequences underlying tissue- or stage-specific expression. Alternatively, this lack of conservation may reflect a lack of selective pressure on a relatively unimportant part of the gene sequence. Exon 2 appears to have been deleted from N116.

Among the three coding exons (exons 3, 4, and 5) displayed in Fig. 2, exon 3 shows the greatest degree of divergence. The first approximately 36 nucleotides of exon three comprise the rest of the 5'-untranslated region, and it is these sequences that were used to generate the transcript-specific probes (see...
### A. NUCLEOTIDES

| EXON 1 | EXON 2 | EXON 3 |
|--------|--------|--------|
| N116   | AGTC TT GACGAGCTTCACG GT... | NT116 | AGTC TT GACGAGCTTCACG GT... |
| N117   | TCTTTCTTCCACAGACTTCTCAAG GT... | N118 | TCTTTCTTCCACAGACTTCTCAAG GT... |
| N119   | TTTGTCCTTCTACAGACTTCTCAAG GT... | N120 | TTTGTCCTTCTACAGACTTCTCAAG GT... |
| N121   | CTCTCTTCTTCTACAGACTTCTCAAG GT... | N122 | CTCTCTTCTTCTACAGACTTCTCAAG GT... |
| N123   | N116   | N117   | N118   | N119   | N120   | N121   | N122   |

**INTRON 1**

**EXON 4**

| Panel A | Panel B |
|---------|---------|
| **INTRON 3** | **EXON 5** |
| N116 | ATGC TCT CCA GAT GTG ATT GCC | N117 | GAG GAC TCT CTG ATG |
| N118 | TTT GTC GTC CAC CCC AAA GAA TCC TGC TTT GTG AAA GGG ACG ATC CAA AAT AAC GAA GCA GGC AAG GTC ACT GTC AGA ACT GCA GGT GGA AGG |
| N119 | N116 | N118 | N117 | N118 | N119 | N120 | N121 | N122 | N123 |

**INTRON 4**

| Panel A | Panel B |
|---------|---------|
| **EXON 6** | **EXON 7** |
| N116 | GAG GTG GTG TTG GCC TAC CGA GGC AAA |
| N117 | AAG GCG CAG GAG GCC CCT CCA CAC ATC TTC TCC ATG |
| N118 | TCTACAGCAGCTTCTTCTCAAG GT... |
| N119 | TCTACAGCAGCTTCTTCTCAAG GT... |
| N120 | TCTACAGCAGCTTCTTCTCAAG GT... |

**FIG. 2.** Comparison of the nucleotide and amino acid sequences of the MHC genes. Panel A, the nucleotide sequences of the seven MHC genes. The genes are arranged in the 5' to 3' direction and aligned to maximize homologies. The exons and introns are marked. N116, N118, and N122 (5, 6, 43) as well as N101, N124, and N125 (44) are shown for comparison. Dashes denote conserved nucleotides between the clone indicated and N116. Gaps within introns indicate possible insertions or deletions. The entire sequences of the 5' untranslated exons are shown for clarity, i.e. no dashes are used. The 5' untranslated sequences used to generate transcript-specific probes are underlined. Only one probe was generated for the N101/N124 pair; the sequence from N101 was used. The same region from N124 is underlined with a dashed line. The translational start site is indicated (M). Panel B, the amino acid sequence of the seven MHC genes. The exon boundaries are indicated. Dashes indicate conserved amino acids with N116 as the standard for comparison.

### B. AMINO ACIDS

| EXON 8 | EXON 9 |
|--------|--------|
| N116 | N117 | N118 | N119 | N120 | N121 | N122 | N123 |

Fig. 2. Comparison of the nucleotide and amino acid sequences of the MHC genes. Panel A, the nucleotide sequences of the seven MHC genes. The genes are arranged in the 5' to 3' direction and aligned to maximize homologies. The exons and introns are marked. N116, N118, and N122 (5, 6, 43) as well as N101, N124, and N125 (44) are shown for comparison. Dashes denote conserved nucleotides between the clone indicated and N116. Gaps within introns indicate possible insertions or deletions. The entire sequences of the 5' untranslated exons are shown for clarity, i.e. no dashes are used. The 5' untranslated sequences used to generate transcript-specific probes are underlined. Only one probe was generated for the N101/N124 pair; the sequence from N101 was used. The same region from N124 is underlined with a dashed line. The translational start site is indicated (M). Panel B, the amino acid sequence of the seven MHC genes. The exon boundaries are indicated. Dashes indicate conserved amino acids with N116 as the standard for comparison.
the legend to Fig. 2). However, the nucleotides surrounding the translational initiation site are highly conserved and agree reasonably well with the consensus sequence for initiation: A/GCCATGG (49–51). The remainder of the exon codes for the amino terminus of the MHC protein. Most of the nucleotide changes in the translated region occur at the silent site (third position) in the codons and result in the conservation of the amino acid. The other nucleotide changes result in only 12 amino acid substitutions, with the majority (eight) being localized in the latter half of the exon. Almost all of the amino acid changes are conservative. One exception occurs at amino acid position 4, where the proline of N116 is replaced by serine in N101, N125, and N127 and is deleted in N118.

Exon 4 shows a very high degree of conservation in the nucleotide and amino acid sequences. Again, most of the nucleotide differences present in the subgroups occur at the silent site in the codons. The only amino acid substitution occurs at amino acid position 78. The substitution is a conservative one.

Finally, exon 5 is completely conserved, except for two silent site changes in N124. This exon is known to contain part of the ATP-binding domain (43, 52, 53), and the almost absolute conservation probably reflects a localized correction mechanism which has occurred to maintain sequences critical for the proteins' function.

The Intron-The first two introns occur within the 5'-untranslated regions of the genes, with the exception of N116 where only one intron is present. We have determined the entire sequence of these introns (except for N101 and N122); only the canonical intron junction sequences and intron lengths are shown in Fig. 2 because of the introns' sizes. Very little sequence conservation is seen within the first two introns; however, the homologies that could be detected are shown in Fig. 6 and Table I.

The sequence conservation of intron 3 strongly emphasizes the aforementioned subgroups: N101/N122, N118/N125, N124/N127, and N116. Numerous insertions/deletions are evident among the four subgroups, but within any particular subgroup, very few differences can be observed.

Intron 4 shows a high degree of conservation among all seven genes. To date, this is the only MHC intron where we have noted this high sequence conservation. Whether the intron serves a particular function or is conserved because of its proximity to the "corrected" fifth exon referred to above is unknown presently.

It is interesting to note that within the coding or translated region the three intron positions (introns 3, 4, and 5) are conserved among the seven genes; these positions are also conserved in all vertebrate MHCs for which sequence information is available (43, 54, 55).

Transcriptional Pattern of N127—The close homologies apparent in the members of the subset are striking and raise the possibility that some of the genes do not encode unique polypeptides, but rather are merely alleles of one another or reflect the polymorphisms known to be present in the MHC gene family (56). In order to explore these possibilities, we have expanded upon our previous transcriptional analyses, which were carried out on N116, N118 (43), and N127 (5) using various methodologies (e.g., N127's transcriptional pattern was determined by S1 analysis (5)). We have shown previously that the 5'-untranslated regions of some of the genes have diverged such that transcript-specific probes can be prepared and used to determine in what tissue and at what developmental stage a particular gene is being expressed (43, 44). Fig. 2 shows a comparison of this region for N127 and five of the subset's members (the analogous region for N122 has not been characterized). The sequences used to generate the various transcript-specific probes are underlined.

The oligonucleotide probe for N127 was end-labeled with T4 polynucleotide kinase and 32p-ATP and hybridized first, to a "Rot blot" containing mRNA isolated from embryonic, neonatal, and adult muscles. The results are shown in Fig. 3A. The probe hybridized strongly to RNA derived from 18-day in vivo cardiac muscle (CAR, lane 6), adult gastrocnemius (GAS, lane 15), and adult posterior latissimus dorsi (PLD, lane 16). The probe also hybridized to mRNA from other muscle tissues; however, the level of hybridization was at or near background levels. Therefore, we were unable to discern at this point if, in fact, N127 was being expressed only in 18-day in vivo cardiac, adult PLD, and adult gastrocnemius muscles. In addition, the hybridization to the 18-day cardiac muscle conflicted with our previous data for this gene's transcript (5).

To confirm these results and resolve the conflicts between these sets of data, a primer extension experiment, which is more sensitive than the Rot blot, was performed. The radioactively labeled oligonucleotide probe was hybridized to the mRNA samples and extended to the 5' ends of the mRNAs using reverse transcriptase. As is the case for the Rot blot, the results define where the homologous transcripts are expressed but, in addition, reveal the length of the transcript and thus are diagnostic only for that particular transcript. In addition, hybridization conditions can be adjusted such that the reactions can be driven to completion using excess probes and can, as a consequence, reveal very low transcript levels. The data are shown in Fig. 3B.

Only four samples had extended products, all of which were 92 bases in length. The positive RNAs were neonatal gastrocnemius (lane 9), neonatal PLD (lane 10), adult gastrocnemius (lane 15), and adult PLD (lane 16). It should be noted that the signals in lanes 9 and 10 were not detectable during a normal time of exposure for the autoradiogram (10 h, −70°C using DuPont Cronex screens and Kodak XRP-5 film). The data in these lanes are a result of a 96-h exposure. Lane 6, which corresponds to the 18-day in vivo cardiac RNA, had no extended product and showed no signal, even when exposures of >7 days were carried out. However, there was a high level of background throughout the entire lane, indicating that the hybridization observed for this RNA in the Rot blot experiment was due to nonspecific binding of the oligonucleotide probe. Alternatively, the transcript may be present in very low abundance in cardiac tissue, and the extended product is masked by the high "noise" that this RNA sample generated.

To confirm that N127 is expressed only in the gastrocnemius and PLD muscles, the end-labeled oligonucleotide probe was annealed in excess to the mRNAs under hybridization conditions such that the reactions were driven to completion. The annealed products were treated with S1 nuclease and electrophoresed on a 12% polyacrylamide 8 M urea gel to detect which RNAs are able to hybridize to, and subsequently protect, the oligonucleotide probe. Previously in this type of analysis, a signal was not detectable in the neonatal muscles (lanes 9 and 10). The data are shown in Fig. 3C and are unambiguous; N127 is expressed during the neonatal and adult stages in the gastrocnemius and PLD musculature. These results extend and confirm the data concerning the expression of N127 (5).

Comparison of the Genes' Transcription—As mentioned above, the transcriptional patterns of other genes within the subset of seven have been determined previously using transcript-specific probes (5, 8, 43) and various methodologies. We thought it useful, based on the above analyses for N127,
Table I

Homologous sequence elements in the 5' end of the MHC genes

The table lists the short, conserved regions depicted in Fig. 6. The residue numbers indicate the location of the sequence relative to the cap site (+1), except for N101. The cap site for N101 has not been determined therefore, the residue numbers indicate the location of the sequence relative to the beginning of the sequenced 5' region of N101. Sequences which show 100% homology are listed on one line; if mismatches are present, the mismatched nucleotides are shown above and below the line and correspond to the first and second genes, respectively. Dashes (-) indicate gaps; D, direct sequence; I, inverted sequence; DR, direct repeat; IR, inverted repeat.

* *, **, *** correspond to longer sequences which are listed below.

| Number | Type | Gene | Residue From | Residue To | Gene | Residue From | Residue To | Homology | Sequence |
|--------|------|------|--------------|------------|------|--------------|------------|----------|----------|
| 1      | IR   | N101 | 37           | 51         | N101 | 820          | 806        | 100.0    | AGCATCTGAGAAGGC |
| 2      | D    | N101 | 132          | 181        | N101 | 780          | 930        | 87.4     | *         |
| 3      | D    | N101 | 192          | 285        | N101 | 917          | 1009       | 80.9     | **        |
| 4      | D    | N101 | 692          | 835        | N101 | 1324         | 434        | 77.4     | ***       |
| 5      | D    | N101 | 643          | 657        | N118 | -858         | -844       | 100.0    | CTTCCAGTAGGATAC |
| 6      | D    | N116 | -596         | -578       | N116 | 2176         | 2194       | 89.5     | ATTTGATGGATAGCA |
| 7      | IR   | N116 | 197          | 221        | N116 | 221          | 197        | 88.0     | AATTTATCGTGAGTGAGATAAAAAT |
| 8      | I    | N116 | 408          | 428        | N124 | 1358         | 1338       | 85.7     | TCTCTGTAAGACAAAGAAATT |
| 9      | I    | N116 | 422          | 439        | N116 | 441          | 423        | 89.5     | GAAATGGTCTGTACAGT |
| 10     | IR   | N127 | -1106        | -1071      | N127 | -1071        | -1106      | 83.3     | AGCACTCTGAGAAGGC |
| 11     | D    | N127 | -243         | -196       | N118 | -199         | -153       | 77.1     | TCCTCTGAGAGATATACT |
| 12     | D    | N127 | -186         | -161       | N118 | -155         | -129       | 76.9     | AGCACTCTGAGAAGGATATATATATC |
| 13     | DR   | N127 | -112         | -101       | N127 | -95          | -70        | 88.5     | GTCCTCGTCTGTTGAGAGATATATAT |
| 14     | D    | N127 | -125         | -100       | N124 | -106         | -82        | 84.6     | TTCCGAGAATGGTCTGTACAGTAAAG |
| 15     | D    | N127 | -95          | -68        | N124 | -122         | -96        | 86.3     | GTCCTCGTCTGTTGAGAGATATATAT |
| 16     | D    | N127 | -95          | -71        | N125 | -111         | -88        | 95.8     | GTCCTCGTCTGTTGAGAGATATATAT |
| 17     | DR   | N127 | 293          | 317        | N127 | 301          | 326        | 76.9     | TCGTCTGAGAATGGTCTGTTGAGAGATATATAT |
| 18     | I    | N127 | -1603        | -1587      | N124 | -1090        | -1106      | 94.4     | TCGTCTGAGAATGGTCTGTTGAGAGATATATAT |

Myosin Promoters
promoters

16541

to repeat the transcriptional analyses using the same methodologies in order that the transcriptional patterns might be easily compared. In addition, the transcriptional pattern of N101 and N124, which has not been analyzed previously, was determined. Although a transcript-specific probe that could distinguish the two could not be prepared because the genes show a high degree of homology in the 5'-untranslated region (see Fig. 2), an oligonucleotide probe for N101/N124 was made. On the basis of the above analysis with N127, we chose to use a combination of Rot blots and primer extensions to carry out a comparative analysis of the genes' expression.

The data are shown in Fig. 4 and illustrate that, despite the close homologies in the exons which encode the amino termini, the genes appear to be unique and are expressed differentially in a tissue- and developmental stage-specific manner.

Of all the genes, N118 appears to be expressed most promiscuously; it is expressed throughout development in a number of different muscle types, although it is expressed at its highest levels during the in ovo stages of development. Adult expression is quite restricted; low levels are present in the adult cardiac and gizzard muscle. The Rot blot (Fig. 4A) and primer extension analysis (Fig. 4B) are in excellent agreement with one another and confirm that this gene is expressed in both skeletal muscle and, in some stages of development, cardiac muscle.

A partial structural analysis of N125 has been described previously (44). A transcriptional analysis has recently been done (73). The data in Fig. 4 show that N125 is expressed predominantly during neonatal development and is probably largely restricted to the fast fibers, as the highest concentrations are present in the pectoralis and PLD muscles. Expression in the adult muscles tested seems to be restricted to the gastrocnemius. Again, the data presented in the Rot blots and primer extensions are completely consistent.

N116 shows a very restricted pattern of expression. The data confirm and extend our previous Northern analysis of the gene's transcription (5). It is expressed predominantly in the adult fast-white musculature, although a small amount of expression in the pectoralis muscles isolated from the 18-day in ovo and neonatal RNAs can be observed in the primer extension analysis (Fig. 4B, lanes 4 and 8). No detectable signals in the corresponding lanes of the Rot blot are present, indicating that the concentration of the transcripts in these tissues is below this assay's detection level.

We were unable to observe any expression of N101/N124 in the Rot blot. However, low level transcription of the gene(s) was detectable upon prolonged exposure (see legend to Fig. 4) of the autoradiogram resulting from the primer extension analysis. Although the 5'-flanking region of N101 has not been determined, the extended product for N124 should be 93 bases (Figs. 2 and 4B), and this is the size of the fragment observed in this assay's detection level.

A Rot blot of N127 is shown here for comparative purposes: it simply confirms the data present in Fig. 3 and illustrates that the highest concentration of transcripts is present in the adult gastrocnemius and in the adult PLD muscles. It should be noted, however, that in the primer extension analysis for this clone (Fig. 4B) in which all lanes in the autoradiogram were developed for an equal period of time (10 h), no detectable signal was present in the neonatal tissues (lanes 9 and 10). This simply underscores the very low levels of transcripts present in these tissues, and extensive development times are needed to observe even a faint signal in these lanes.
Fig. 3. Transcriptional pattern of N127. Panel A, the results of a Rot blot experiment using the transcript-specific probe of N127. 500,000 cpm/ml of the probe was hybridized to the filter in 6 x SSC, 0.1% sodium dodecyl sulfate for 20 h at 33°C. The filter was washed four times in 5 x SSC, 0.1% sodium dodecyl sulfate, and 0.1% sodium pyrophosphate for 15 min at 37°C, and an autoradiogram was developed. The muscle sources for the mRNAs are abbreviated as follows: BRT, breast; CAR, cardiac; LEG, leg; SP, superficial pectoralis; DP, deep pectoralis; GAS, gastrocnemius; GIZ, gizzard, ALD, anterior latissimus dorsi; PLD, posterior latissimus dorsi. The neonates were 9 days post-hatched, while the adults were 6 months post-hatched or older. Panel B, the results of a primer extension experiment using the probe for N127. The bands in each lane correspond to the extended product of 92 nucleotides. No other bands were detected in the other RNAs. Panel C, the results of an S1 analysis. The band in each lane corresponds to the fully protected probe obtained following the S1 nuclease digestion. The RNAs present in lanes 9 and 10 were present at three times the amount of the other RNAs. bp, base pairs.

Fig. 4. Comparative transcriptional analysis of the genes. Panel A, Rot blots. The Rot blots were carried out essentially as described in Fig. 3. Panel B, primer extension analyses were carried out as described previously (43); equal amounts of RNA (15 μg) were used in each of the lanes. The autoradiograms for N118, N125, N116, and N127 were the results of a 10-h exposure at -70°C with DuPont Cronex screens and Kodak XRP-5 film. The exposure for N101/ N124 was carried out for 72 h using identical conditions. The organization of the 5'-untranslated regions is shown schematically on the right. The numbers adjacent to the boxes on the right side of the diagram correspond to the size of the extended primer.
FIG. 5. Nucleotide sequences of the promoter regions of the genes. The sequences have been aligned to maximize the homologies. The sequences upstream of -240 have been omitted in N127, N125, and N118 to emphasize the homologies between N116 and N124. No other homologies existed upstream of -240 when comparing all five genes, except those shown for N116 and N124. Areas showing high conservation are boxed. The TATA and possible CAAT boxes are marked. The transcriptional start sites are indicated by +1, e.g. in N116, it is an A; in N118, it is a T; and in N127, N124, and N125, it is a G. The first exon/intron junction is marked by a vertical line.

FIG. 6. The 5' ends of the MHC genes. The genes are drawn to scale. The coding and untranslated exons are represented as open and closed boxes, respectively. The TATA and CAAT boxes are represented as vertical lines. Sequences homologous to the chicken repetitive sequence (CRI) are shown ( ). The conserved regions are represented as horizontal lines above each clone. Each region is numbered and the sequence shown in Table I. A large tract of a single nucleotide is denoted by an arrow; the number of nucleotides within the tract is also indicated.

is seen throughout development in various muscle tissues, has only one.

DISCUSSION

The results presented here show that the organization of the seven chicken MHC genes at the 5' ends is conserved. In the coding region, all of the genes are interrupted by introns at the same position, while in the 5'-untranslated regions, all of the MHC genes studied to date are interrupted by introns at highly conserved positions with the exception of N116. It would appear that the second untranslated exon has been deleted in N116; alternatively, this exon was inserted into a gene(s) that subsequently underwent a duplication(s). We favor the first hypothesis, because other MHC genes from different species also have two exons in the 5'-untranslated region (11, 59). The 5'-untranslated regions of the genes have diverged and have proved to be useful for generating transcript-specific probes. The data show clearly that (with the possible exception of N101) the genes are not polymorphic variants or alleles of one another because each exhibits a unique pattern of transcription.

All of the gene's promoters analyzed contain a CAAT and TATA box at the expected locations, and the sequences in and around these regions are highly conserved. However, the promoter sequences of the cardiac α-MHC genes from rat (59) and rabbit (60) are compared with the five chicken promoter regions shown, no homologies were observed. Furthermore, a 400-base region, which is 80% homologous between the rat and rabbit MHC 5' sequences (59), is not found in any of our chicken genes. Recently, a rat embryonic skeletal muscle MHC promoter was determined (55). When the sequence from -303 to -1 of the rat promoter was compared with the chicken promoters, little sequence conservation could be observed. The highest degree of homology
pressed genes (66). The exact role of these regions in the DNA (65) which may be sites of protein binding or areas Figure 6). These areas presumably represent bends or kinks in the DNA. For example, regions 20 and 20a contain alternating transcriptional control (61-66) are represented in these homologies. For example, regions 20 and 20a contain alternating transcriptional control (61-66) are represented in these homologies.

The 5' flanking regions of the genes in terms of short homologies. Fig. 6 schematically depicts the organization of the genes' 5' ends and the locations of short regions (>15 bases) of homology (>75%) which are shared between two of the genes. The nucleotides present in each of the homologies are shown in Table I. Various sequences which have been hypothesized to function in different aspects of transcriptional control (61-66) are represented in these homologies. For example, regions 20 and 20a contain alternating purine/pyrimidine residues that may represent Z-DNA (62), although what role (if any) this structure plays in regulation remains to be determined (62-64). Adenine- and thymidine-rich regions are also present in the 5' ends of the genes (see Fig. 6). These areas presumably represent bends or kinks in the DNA (65) which may be sites of protein binding or areas of disfavored nucleosome formation. In yeast, such sites represent upstream promoter elements of constitutively expressed genes (66). The exact role of these regions in the MHC genes remains to be determined.

The 5' ends of the genes were searched for proposed muscle-specific sequences (67-69). Only one was detected at 100% homology, CATTCTCT; this sequence is found in a number of muscle-specific genes (e.g. actin, troponin I, troponin T, and myosin light chain) from a variety of species. In the MHC genes, the sequence is located at -193 in N116, -418 in N125, -487 in N118, and -610 in N127. The sequence is also located at +70 in N124, +103 in N108, +373 in N125, +433 in N127, and +2052 in N116. Therefore, if this sequence is involved in the transcriptional control of the contractile protein genes, it seems that its effect is not dependent upon an absolute location relative to the transcriptional start site.

Other regions besides the 5'-flanking and untranslated regions may also be involved in the expression of the MHC genes. Intragenic regions have been implicated in the regulation of other genes (40, 70). However, the short regions of homology catalogued herein are logical candidates for cis-acting regulatory elements (71, 72) and provide a starting point for the functional elucidation of sequences that are involved in the controlled expression of the MHC genes.

REFERENCES

1. Robbins, J., Freyer, G. A., Chiholm, D., and Gilliam, T. C. (1982) J. Biol. Chem. 257, 549-556
2. Nguyen, H. T., Gubits, R. M., Wydro, R. M., and Nadal-Ginard, B. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 5230-5234
3. Kanicky, C. J., Umeda, P. K., Sinha, A. M., Elrige, M., Tong, S. W., Zak, R., Jakovic, S., and Rabinowitz, M. (1983) J. Biol. Chem. 258, 5196-5205
4. Wydro, R. M., Nguyen, H. T., Gubits, R. M., and Nadal-Ginard, B. (1983) J. Biol. Chem. 258, 670-678
5. Robbins, J., Horan, T., Glick, J., and Kropp, K. (1986) J. Biol. Chem. 261, 6606-6612
6. Robbins, J., Kropp, K., Glick, J., Horan, T. P., and Valente, N. (1986) UCLA Symp. Mol. Cell. Biol. New Ser. 29, 375-385
7. Whalen, R. G., Sell, S. M., Butler-Browne, G. S., Schwartz, K., Lozovoret, V., and Finset-Hartarmac, I. (1981) Nature 292, 805-809
8. Mahdavi, V., Periasamy, M., and Nadal-Ginard, B. (1982) Nature 297, 659-664
9. Lampre, A.-M., Nadal-Ginard, B., and Mahdavi, V. (1984) J. Biol. Chem. 259, 6437-6446
10. Izumo, S., Nadal-Ginard, B., and Mahdavi, V. (1986) Science 231, 597-600
11. Mahdavi, V., Streher, E. P., Periasamy, M., Wieczorek, D., Izumo, S., Grond, S., Streher, M. A., and Nadal-Ginard, B. (1986) UCLA Symp. Mol. Cell. Biol. New Ser. 29, 345-361
12. Shih, I. F. Y., McGee, A. L., and Hale, P. T. (1978) J. Mol. Cell. Cardiol. 10, 1053-1076
13. Chizzonite, R. A., Everett, A. W., Clark, W. A., Jakovic, S., Rabinowitz, M., and Zak, R. (1982) J. Biol. Chem. 257, 2055-2065
14. Lombé, A.-M., Schwartz, K., d’Albis, A., Lacombe, G., Van Thiem, N., and Swynghedauw, B. (1979) Nature 282, 105-107
15. Butler-Browne, G. S., Bugaisky, L. B., Cuinoud, S., Schwartz, K., and Whalen, R. G. (1982) Nature 299, 830-833
16. Gauthier, G. F., Burke, R. E., Lowey, S., and Hobbs, A. B. (1983) J. Cell Biol. 97, 756-771
17. Matautia, R., Spector, D., and Strohman, R. C. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 1122-1125
18. Strohman, R. C., Moss, P. S., Micou-Eastwood, J., Spector, D., Przybyla, A., and Paterson, B. (1977) Cell 10, 265-273
19. Medford, R. M., Wydro, R. M., Nguyen, H. T., and Nadal-Ginard, B. (1986) Proc. Natl. Acad. Sci. U. S. A. 77, 5749-5753
20. Benoff, S., and Nadal-Ginard, B. (1980) J. Mol. Biol. 140, 283-298
21. Caravatti, M., Minty, A., Robert, B., Montarras, D., Weydert, A., Cohen, A., Daubus, P., and Buckingham, M. (1982) J. Mol. Biol. 160, 59-76
22. Medford, R. M., Nguyen, H. T., and Nadal-Ginard, B. (1983) J. Biol. Chem. 258, 11063-11073
23. Mroczkowski, B., McCarthy, T. L., Zetza, D. J., Bragg, P. W., and Heywood, S. M. (1984) Exp. Biol. Med. 9, 277-283
24. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383
25. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980) Science 209, 1406-1414
26. Grosschedl, R., and Birnstiel, M. L. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 1432-1436
27. Ghosh, P. K., Lehewitz, P., Frisate, R. J., and Gluzman, Y. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 190-194
28. Benist, C., and Chambon, P. (1981) Nature 290, 304-310
29. Benoist, C., O’Hare, K., Breathnach, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142
30. Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J., and Weissmann, C. (1983) Cell 32, 695-706
31. McKnight, S. L., and Kingsbury, R. C. (1982) Science 217, 316-325
32. McKnight, S. L., Gavis, E. R., Kingsbury, R. C., and Axel, R. (1981) Cell 25, 395-398
33. McKnight, S. L., Kingsbury, R. C., Spence, A., and Smith, M. (1984) Cell 37, 253-262
34. Gelinas, R., Endlich, B., Pfeiffer, C., Yagi, M., and Stamatoyanopoulos, G. (1985) Nature 313, 323-325
35. Charnay, P., Mellon, P., and Maniatis, T. (1985) Mol. Cell. Biol. 5, 1495-1511
36. Jones, K. A., Yamamoto, K. R., and Tjian, R. (1985) Cell 42, 559-572
37. Dyman, W. S., and Tjian, R. (1985) Nature 316, 774-778
38. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gauth, M. P., and Chambon, P. (1983) Nucleic Acids Res. 9, 6047-6058
39. Gluzman, Y., and Shenk, T. (eds) (1983) in Enhancers and
Myosin Promoters

1. Gillies, S. D., Morrison, S. L., Oi, V. T., and Tonegawa, S. (1983) Cell 33, 717-728
2. Chandler, V. L., Maler, B. A., and Yamamoto, K. R. (1983) Cell 33, 489-499
3. Gulick, J., Kropp, K., and Robbins, J. (1985) J. Biol. Chem. 260, 14513-14520
4. Kropp, K., Gulick, J., and Robbins, J. (1986) J. Biol. Chem. 261, 6613-6618
5. Lele, R. M. K., McClure, B. A., and Houchins, J. P. (1985) Plasmid 13, 31-40
6. Amersham Corp. (1983) M13 Cloning and Sequencing Handbook, Amersham Corp., Arlington Heights, IL
7. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
8. Queen, C., and Korn, L. J. (1984) Nucleic Acids Res. 12, 581-599
9. Kozak, M. (1984) Nature 308, 241-246
10. Kozak, M. (1986) Cell 44, 283-292
11. Tong, S. W. and Elzinga, M. (1983) J. Biol. Chem. 258, 13100-13110
12. Okamoto, Y., and Yount, R. G. (1985) Proc. Natl. Acad. Sci. U. S. A. 85, 1575-1579
13. Streher, E. E., Mahdavi, V., Periasamy, M., and Nadal-Ginard, B. (1985) J. Biol. Chem. 260, 468-471
14. Streher, E. E., Streher-Page, M.-A., Perriard, J.-C., Periasamy, M., and Nadal-Ginard, B. (1986) J. Mol. Biol. 190, 291-317
15. Gillam, T. C. (1982) Ph.D. dissertation, University of Missouri-Columbia
16. Zweig, S. E. (1981) J. Biol. Chem. 256, 11847-11853
17. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560
18. Mahdavi, V., Chambers, A. P., and Nadal-Ginard, B. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 2626-2630
19. Umeda, P. K., Levin, J. E., Sinha, A. M., Cribbs, L. L., Darling, D. S., Ende, D. J., Hau, H.-J., Dixon, F., and Jakovic, S. (1986) UCLA Symp. Mol. Cell. Biol. New Ser. 29, 809-823
20. Wang, A. H.-J., Quigley, G. J., Kopsak, F. J., Crawford, J. L., van Boom, J. H., van der Marsel, G., and Rich, A. (1979) Nature 282, 680-686
21. Nordheim, A., and Rich, A. (1983) Nature 303, 674-679
22. Marx, J. L. (1985) Science 230, 794-796
23. Nordheim, A., Pardue, M. L., Lafer, E. M., Moller, A., Stollar, B. D., and Rich, A. (1981) Nature 284, 417-422
24. Koo, H.-S., Wu, H.-M., and Crothers, D. M. (1986) Nature 320, 501-506
25. Struhl, K. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 8419-8423
26. Daubas, P., Robert, B., Garner, L., and Buckingham, M. (1985) Nucleic Acids Res. 13, 4622-4643
27. Minty, A., and Kedes, L. (1986) Mol. Cell. Biol. 6, 2125-2136
28. Nikovits, W., Jr., Kuncio, G., and Ordahl, C. P. (1986) Nucleic Acids Res. 14, 3377-3390
29. Charnay, P., Treisman, R., Mellen, P., Choa, M., Axel, R., and Maniatis, T. (1984) Cell 38, 251-263
30. Britten, R. J., and Davidson, E. H. (1969) Science 165, 349-353
31. Davidson, E. H., Jacobs, H. T., and Britten, R. J. (1983) Nature 301, 468-470
32. Gulick, J., Kropp, K. E., and Robbins, J. (1987) Eur. J. Biochem., in press