Intergenic Bidirectional Promoter and Cooperative Regulation of the IIx and IIb MHC Genes in Fast Skeletal Muscle

C. Rinaldi, F. Haddad, P.W. Bodell, A. X. Qin, W. Jiang, and K. M. Baldwin

University of California Irvine, Physiology and Biophysics Dept, Irvine, CA 92697

Running Title: Cooperative Regulation of the IIx and IIb MHC Genes

Key words: gene transcription, pre-mRNA, natural antisense RNA, comparative genomics.

Corresponding Author

Kenneth M. Baldwin

kmbaldwi@uci.edu
ABSTRACT

This study investigated the dynamic regulation of IIx-IIb MHC genes in the fast white medial gastrocnemius (WMG) muscle in response to intermittent resistance training exercise (RE), a model associated with a rapid shift from IIb to IIx expression (11). We investigated the effect of four days of RE on the transcriptional activity across the skeletal MHC gene locus in the WMG in female Sprague Dawley rats. Our results show that RE resulted in significant shifts from IIb to IIx observed at both the pre-mRNA and mRNA levels. An antisense RNA (xII NAT) was detected in the intergenic region between IIx and IIb, extending across the entire IIx gene and into its promoter. The expression of the xII NAT was positively correlated with IIb pre-mRNA (R=+0.8), and negatively correlated with IIx pre-mRNA (R=-0.8). Transcription mapping of the IIx-IIb intergenic region revealed the generation of sense IIb and xII NATs from a single promoter region. This bidirectional promoter is highly conserved among species and contains several regulatory elements that may be implicated in its regulation. These results suggest that the IIx and the IIb genes are physically and functionally linked via the bidirectional promoter. In order for the IIx MHC gene to be regulated, a feedback mechanism from the intergenic xII NAT is needed. In conclusion, the intergenic bidirectional promoter generating antisense RNA appears to be essential for the coordinated regulation of the skeletal muscle MHC genes during dynamic phenotype shifts.
INTRODUCTION

Skeletal muscle tissue is able to adapt to different types of stimuli such as neuronal, mechanical and hormonal interventions (8). This plasticity is made possible, in part, by the existence of several isoforms of myosin heavy chain (MHC), the molecular motor of muscle contraction. At least 8 MHC isoforms are expressed in mammalian striated muscle (skeletal and cardiac) including two developmental isoforms (embryonic and neonatal), one slow MHC (type I or β cardiac), one cardiac isoform (α MHC), three fast type II isoforms (IIa, IIx, and IIb), and one specialized isoform (the extraocular MHC). These MHC isoforms are the products of distinct genes that are clustered in multigenic complexes on 2 chromosomes. The skeletal MHC gene cluster resides on chromosome 10 in the rat (10q24) (17p13 in human, 11q35.0 cM in mouse), spanning ~420kb. These MHC genes are linked in a head to tail fashion in the following order: embryonic, IIa, IIx, IIb, neonatal and extraocular MHC (28, 29, 41, 49, 51). The cardiac MHC gene cluster resides on chromosome 15 in the rat (15p13) (14q12 in human and 14q20.0 cM in mouse), spanning ~50 kb; and consists of the β (slow type I MHC) approximately 4.5 kb upstream of the αMHC gene (16, 34, 39). Different combinations of MHC isoforms may occur in the same fiber, and the MHC isoform profile is a major determinant of muscle fiber functional properties, such as the speed of contraction, power output, and fatigue resistance. In adult rodents and small body size mammal skeletal muscle, all four types of protein isoforms, MHC-I, IIa, IIx and IIb, are expressed (40). In humans, even though the MHC IIb gene has been identified (41), the general belief, based on current detection methods, is that the IIb MHC gene is not expressed in human muscle fibers (31, 42, 50). However, IIb isoform was abundantly detected in a subset of human masseter muscle fibers, and was rarely detected in limb muscle and in few external oblique fibers, but these detections occurred only at the mRNA level (24).
The expression of MHC isoforms is both developmentally regulated and highly responsive to mechanical, neuronal or hormonal stimuli (6, 7, 10, 40). Based on the type of intervention and the starting phenotype, muscles can shift towards a slower or a faster contractile profile (6). MHC isoform regulation occurs in a muscle type specific fashion in response to various stimuli. For example, the IIx MHC expression is upregulated in a slow muscle fiber type in response to unloading or inactivity (18, 23, 25, 45); whereas, the same IIx MHC expression is also upregulated in a fast muscle in response to overloading and increased activity such as resistance training (11, 22). The upregulation of the IIx gene in response to opposing stimuli in two different muscle types is intriguing and is thought to occur at the transcriptional/pretranslational level (37, 38). The exact mechanism behind this regulation is not clear but it may be the result of both muscle type specific factors acting on the genetic apparatus, as well as cross talk between the IIx gene and its adjacent genes (IIa and IIb).

The MHC gene family organization, tandem linkage, individual gene order and spacing have been conserved through million of years of evolution; and thus these physical characteristics are believed to play a key role in their coordinated regulation and dynamic switching that occurs between the MHC isoforms during development and in response to altered physiological conditions (13, 17, 20, 26, 50). In fact, in recent years, our studies examining MHC gene switching uncovered that the intergenic DNA physically linking the two MHC genes is transcriptionally active and thus, may provide a functional link between these tandem MHC genes, thus leading to their coordinated regulation. For example, the intergenic (IG) space between cardiac β and α MHC genes is transcriptionally active on both strands thus exhibiting some inherent bidirectional transcriptional activity (20). In a normal control rodent heart, as well as in response to thyroid hormone (T3), the intergenic activity is turned on leading to an increase in both sense and antisense IG transcription. The sense IG RNA is correlated to an increased αMHC gene transcription, whereas the antisense transcription is
carried out through the entire βMHC gene into its promoter whereby it is believed to interfere with its transcription via a mechanism likely involving epigenetic processes (20, 21, 38). The intergenic transcriptional activity is thought to play a critical role in the coordinated antithetical regulation of these two cardiac MHC genes.

In a more recent report we show that the intergenic region between the IIa and IIx genes also exhibits bidirectional transcriptional activity in slow muscles such as the soleus and vastus intermedius (VI) (38). This intergenic transcription is thought to be a player in the coordinated regulation of IIa-IIx MHC in response to unloading and inactivity (38).

The goal of this current study was to further examine the MHC isoform shifts in the fast white medial gastrocnemius (WMG) muscle in response to increased activity such as high resistance training exercise (RE). Under normal control conditions, the adult rodent WMG MHC isoform profile is composed of ~80-90% IIb MHC and 10-20% IIx MHC, and this composition is found at both the protein and mRNA level of expression. In rodents, RE can induce rapid shifts in MHC isoform from IIb to IIx, during which the IIb is downregulated whereas the IIx is upregulated (11, 22). These IIb→IIx shifts occur rapidly and can be detected at the mRNA level after only two days of RE (11). The IIx (Myh1) and IIb (Myh4) MHC genes are located in tandem on the skeletal MHC gene locus, and are separated by ~14kb of intergenic DNA in the rat (NCBI rat genomic database, accession # NW_047334). Consequently, one goal of this study was to determine if the intergenic DNA between the IIx and IIb genes is transcriptionally active on both strands, and if this intergenic locus serves as a focal point in their coordinated regulation in WMG in response to RE. Thus, we tested the working hypothesis that the same mechanism of coordinated regulation, described for the cardiac MHC genes and for the skeletal muscle MHC gene IIa and IIx, is also involved in the
switching of the MHC genes IIx and IIb in the WMG. Type IIx and IIb pre-mRNA and mature mRNA were analyzed as markers of transcriptional and pre-translational events respectively. In addition, sense and antisense RNA were analyzed along the IIx-IIb MHC gene locus including the intergenic DNA. We investigated the effect of four days of RE on these MHC RNA markers. The second hypothesis tested was that the co-regulated modulation of the IIb sense and IIx antisense transcripts is linked to the presence of a common bidirectional promoter situated in the IIx-IIb intergenic region. Transcription was mapped through the entire IIx-IIb intergenic region and rapid amplification of cDNA 5’ ends analyses (5’RACE) were conducted in order to identify transcription start sites. The obtained findings corroborate the generation of the sense and antisense transcripts from a single promoter region. Multi-species comparative sequence analyses (in silico) was conducted in order to identify potential transcription factors binding sites which could be involved in the regulation of the bidirectional promoter.

METHODS

ANIMAL MODEL AND EXPERIMENTAL DESIGN.

This study was conducted in conformity with the Guiding Principles in the Care and Use of Animals of the American Physiological Society, and the protocol was approved by the University of California, Irvine, Animal Use Committee. This study used young adult female Sprague-Dawley rats weighing 256 ±5 g (n = 12). Rats were on a 12:12-h light-dark cycle and were allowed access to food and water ad libitum. The rats experienced unilateral leg RE via imposed electrical stimulation (see below) targeting the left gastrocnemius muscle. The experiment lasted for 6 days, RE was administered on day 1, 2, 4, and 5. Day 3 was a rest day. For each RE bout, the rats were anesthetized with ketamine-xylazine-acepromazine (30/4/1 mg/kg). On day 6, i.e., 24 h after the last resistance exercise session, the experiment
was terminated, and the animals were euthanized for tissue collection via an IP injection of Pentosol euthanasia solution (Med-Pharmex) at a dose of 0.4 ml/kg (~160mg/kg Pentobarbitol). The medial gastrocnemius muscle (MG) from both legs was dissected out, and was divided into white MG (WMG), red MG, and a mix MG portions. Muscle pieces were immediately frozen between blocks of dry ice and were stored at -80°C for subsequent RNA extraction and analyses. The contralateral limbs served as an internal control, and data values are therefore designated as “CON” as opposed to RE for the trained leg data.

TRAINING PROTOCOL.
We have used a combined muscle action training protocol (isometric, concentric, eccentric) similar to the one described in Adams et al (2). Each bout of stimulation lasted 3 seconds, the first second was of the isometric type contraction followed by a 1 second concentric contraction, followed by a 1 second eccentric mode contraction. A period of 27 s of rest was allowed between each contraction. Rats performed 5 contractions per set. Five minutes of recovery was allowed between sets. On day 1, the session was limited to five sets, on the subsequent 3 training days, the animals received six sets of contractions during each training session.

RNA ANALYSIS.
Total RNA was extracted from frozen WMG muscle using the Tri Reagent protocol (Molecular Research Center). The extracted RNA was DNase-treated using one unit of RQ1 RNase-free DNase (Promega) per µg of total RNA and was incubated at 37°C for 30 minutes followed by a second RNA extraction using Tri Reagent LS (MRC).

The MHC mRNA isoform distribution was evaluated by reverse transcription (RT) with random primers/oligo dT mix followed by polymerase chain reaction (PCR) with
primers targeting the embryonic, neonatal, I, IIA, IIX, and IIB MHC mRNAs, as described previously (3, 15). In these PCR reactions, each MHC mRNA signal was corrected to an externally added control DNA fragment that was co-amplified with the MHC cDNAs using the same PCR primer pair. This provides a means to correct for any differences in the PCR reaction efficiency and/or pipetting of each PCR reaction. A correction factor was used for each control fragment band on the ethidium bromide-stained gel to account for the staining intensity of the variably sized fragments (224 to 324 bp), as reported previously (15).

While the above RT-PCR can provide information on the MHC mRNA distribution pattern, it does not give information on how each isoform is regulated. Strand-specific RT-PCR was used in order to analyze the expression of specific MHC pre-mRNAs and mRNAs, as well as antisense RNAs that are of opposite orientation to the MHC genes.

Strand specific RT-PCR used the one step-RT-PCR kit from Qiagen. These assays were utilized in the determination of the relative level of expression of pre-mRNA, antisense RNA, and mRNA in a known amount of total RNA, in comparing NC vs. RE WMG. Also, these analyses were carried out in the mapping analyses of RNA expression across the skeletal MHC gene locus. The provided instruction was followed except in the RT step, only the reverse primer was added to target the sense RNA, whereas only the forward primer was added to target the antisense RNA. At the end of the RT, the reaction was heated to 90°C for 15 minute to denature the RT enzyme, then the missing primer was added before initiating the PCR. In these assays, a negative control reaction was carried out to ensure the strand specificity of the product. In the RT step, no primer was added. After denaturing the RT, both the Fwd and Rev PCR primers were added. Any resulting product from this negative control reaction may be indicative of the formation of a cDNA in a primer-independent fashion, or alternatively may be indicative of the presence of genomic DNA in the RNA. Our tests show the absence of any products in the negative control reactions, confirming that any amplified
product corresponded to the target sense or antisense RNA in question (see (19) for more details).

For all the PCR analyses, the primers were designed using DNA-star Primer Select software, in which the target sequence was repeat masked using the Repeat Masker web server at http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker.

5’RACE: RAPID AMPLIFICATION OF THE CDNA ENDS.

These assays were conducted in order to identify the TSS of both the antisense IIx RNA and the sense IIb RNA. We adapted the method described by Invitrogen supplied with the kit (5’RACE). The cDNA was synthesized using Superscript II at 50°C and an RT primer reverse complementary to a region of the target RNA near the 5’ end. The cDNA was RNase H treated and purified using the Qiaquick PCR Purification Kit (Qiagen). The pure cDNA was tailed with dCTP using terminal transferase (New England Biolab). The tailed cDNA was amplified using a gene specific reverse primer that was nested to the RT primer and an adapter primer (5’RACE primer, Invitrogen) (see supplementary table S1 for the RT and gene-specific primer sequence). The PCR products were cloned into pGEMT easy (Promega) and the insert of individual clones was sequenced in order to determine the 5’ end corresponding to the end of the cDNA that is adjacent to the poly dC tail. Sequencing reactions used the ABI Big Dye sequencing kit version 3.1. Sequencing data acquisition was obtained from the UCI sequencing facility using an Applied Biosystem 3700 sequencer.

GENOMIC DNA.

Genomic DNA was extracted from rat tissue using the DNAeasy Tissue Kit (Qiagen). DNA was eluted with water and its concentration was determined by UV absorption at 260 nm (using a factor of 50 µg/ml per OD unit). The obtained genomic DNA was used as template
to test all the PCR primer sets targeting pre-mRNA and antisense RNA to ensure that these primers worked with similar efficiency to each other.

COMPUTATIONAL ANALYSES OF DNA SEQUENCE.
Rat genomic sequence (genomic contig number: NW 047334, location: 10q24) was aligned with human (contig number: NT 010718, location: 17p13), mouse (contig number: NT 096135, location: 11 35cM) and Monkey (Genbank accession #: AC212180) using Mulan comparative genomics site at http://mulan.dcode.org/ (36). The purpose for this comparative analysis was to identify evolutionary conserved sequences at 85% in these promoters and identify conserved transcription factor binding sites in these conserved regions.

STATISTICAL ANALYSES.
Data are reported as means ± SEM. Differences between two groups (CON vs. RE) were analyzed using a paired t-test. Relationships between two variables were assessed using linear regression and correlation analyses (GraphPad Software, Inc.). One way ANOVA followed by Newman Keuls post hoc test was used for the analyses of promoter activity across muscle types. Statistical significance was set at p< 0.05.
RESULTS

MUSCLE WEIGHTS

After 4 days of RE the ratio of MG muscle weight/body weight was increased by 5% in the trained leg relative to the contra-lateral control leg, although this increase was not statistically significant. CON MG muscle was 630±16 mg; RE muscle weight was 662±17 mg.

MHC mRNA ANALYSIS

The MHC mRNA isoforms distribution in the WMG and their shift following four days of RE were assessed through a competitive PCR approach and they are shown in Figure 1. The mRNA of each MHC gene was expressed as the percent of the total MHC mRNA pool. Confirming previous findings (11, 22), only fast MHC isoforms, IIx and IIb, were expressed in both groups of samples. Since other MHC mRNA isoforms were not detected either in CON or in the RE group, they were not reported in the graph. The MHC mRNA distribution in control WMG muscles was represented by 76% of IIb. The acute RE paradigm caused a statistically significant shift in the MHC composition, with a decrease of MHC IIb mRNA to 50% and a consequent increase of MHC IIx mRNA from 24 to 47%. These results were consistent with a previous report on MHC isoform shifts following mechanical loading of fast skeletal muscles (11). Furthermore, these results demonstrate that the RE in these current experiments was effective in inducing MHC isoform shifts which warrant further analyses of RNA across the gene locus.

MHC RNA EXPRESSION IN RESPONSE TO RESISTANCE LOADING.

The IIx and IIb MHC pre-mRNA and mRNA were used as markers of transcriptional and of pre-translational events, respectively. In these analyses, both the mRNA and the pre-mRNA are expressed as arbitrary units (AU) per nanogram (ng) of total RNA, whereas in figure 1,
the MHC mRNA was expressed as relative % of the total MHC mRNA pool. Results in Figure 2 show that for the IIx MHC gene products, both the pre-mRNA and the mature mRNA signals (Figure 2, A and B) were significantly increased in response to RE (P<0.0001). The IIx MHC pre-mRNA was ~ 5 fold higher in RE muscles than in the control group and the IIx MHC mRNA recorded was 6.5 fold greater (Figure 2B). For the IIb MHC gene products, the expression of pre-mRNA and the mRNA was significantly decreased in RE muscles (P<0.001). The RE WMG expressed only 25% and 40% of the IIb pre-mRNA and mRNA, respectively, as compared to control muscle (Figure 2, D and E).

In addition to the sense RNA species for the IIx and IIb MHC genes, the antisense RNA was analyzed in several locations across the IIx MHC gene and in the IIx-IIb intergenic region. The IIx MHC antisense RNA (xII NAT), as measured in the intergenic region (at ~2 kb from the IIx TGA stop codon), was decreased significantly in the RE muscles (RE is 9% of the CON group; P<0.001) (Figure 2C). Interestingly the antisense transcription persisted along the entire IIx gene and into its promoter region 5’ upstream of the IIx gene (see below). No signal of antisense RNA was found in the IIb MHC gene, either in CON or in RE muscles. The antisense transcript corresponding to the IIb MHC gene was targeted at several locations across the IIb gene, including its 3’ flanking region.

The relationship between xII NAT vs IIx MHC pre-mRNA and the relationship between xII NAT vs. IIb MHC pre-mRNA were analyzed by linear regression. As presented in Figure 3, there was a significant inverse correlation between the xII NAT and the IIx MHC pre-mRNA (R= -0.8; P<0.001). In contrast, there was a significant positive correlation between the xII NAT and that of the IIb pre-mRNA (R= +0.8; P<0.0001) (Figure 3). These relationships show that the behavior of the IIx-IIb gene locus transcriptional activity is remarkably similar.
to those described for the cardiac β and αMHC genes (17) as well as for those for the IIa-IIx MHC gene in the soleus muscle (38). The coordinated regulation of the IIx-IIb gene concurs with the previously described models for two linked genes.

**TRANSCRIPTION MAPPING ACROSS THE IIA-IIX-IIB GENES IN WMG MUSCLE.**

The above reported results demonstrate that an antisense IIx RNA is expressed in fast WMG muscle fibers, and its expression is down regulated in response to RE. It was of interest to map the extent of sense and of antisense transcription along the adult skeletal MHC gene locus (IIa-IIx-IIb) in the fast WMG. Based on the rat genomic DNA sequence in NCBI (accession NW_047334, Rattus norvegicus chromosome 10 genomic contig, reference assembly; RGSC v3.4), these genes cover a region of ~90kb of genomic DNA, of which the IIa gene is ~27kb, whereas the IIx and the IIb genes are ~23kb each in size. The IIa and IIx MHC genes are separated by ~2.7kb, whereas the IIx-IIb MHC genes are separated by ~14kb of intergenic space. Within the 14kb intergenic space, there is a gap of ~661 bp of unknown sequence centered at ~-3kb relative to the IIb MHC gene transcription start site (TSS). Several DNA repetitive elements are found on the IIx-IIb intergenic sequence as revealed by using a repeat masker tool (Repeat Masker web server; url: http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). The masking results showed that these genomic repeats make up ~31% of the total IIx-IIb intergenic region (figure 4).

A strand specific RT-PCR based approach was utilized to target and amplify the sense and antisense RNA in the white MG total RNA at sites corresponding to various regions of the IIa-IIx-IIb MHC gene locus. Strand-specific RT-PCR results demonstrated that the expression of antisense (AS) RNA in WMG muscles can be detected across the entire IIx gene, its promoter region consisting of the IIa-IIx intergenic region (2677 bp), and in the
region corresponding to the 3’ end of the IIa gene (see figure 4). Also, AS RNA was detected corresponding to a large segment of the intergenic region between the IIx and IIb MHC genes to within approximately -200 bp relative the IIb MHC gene TSS (+1). No sense RNA was detected in the control WMG across either the IIa gene or the IIa-IIx intergenic region. In a few trained WMG samples, some trace levels of IIa sense RNA was detected. A significant amount of sense RNA was detected across the entire IIx MHC gene, the entire IIb MHC gene, and to ~3kb past the IIx or IIb gene stop codon (Figure 4). Furthermore, these mapping analyses detected no antisense RNA that corresponds either to the IIb gene or to its 3’ flanking sequence. To our surprise, a low level of sense RNA was detected across ~400bp of the IIb MHC proximal promoter. This sense RNA merged with the IIb gene at +1 and it became undistinguishable from the IIb pre-mRNA. This low level of sense RNA corresponding to the IIb proximal promoter could not be attributed to a non specific signal or contamination since no bands were observed in the negative control reactions.

A high resolution mapping of sense and antisense transcriptional activity in the proximal IIb promoter (-1kb to +1kb vs IIb TSS) is shown in figure 4. These mapping analyses suggest that both sense and antisense intergenic (IG) transcriptions are occurring in the proximal IIb promoter. Both of these sense and antisense IG RNA start within 400 bp from the IIb MHC gene TSS. 5’RACE analyses using primer extension and tailing reactions followed by sequencing, revealed the presence of multiple start sites of the antisense RNA that are clustered between -403 to -271 bp relative to the IIb MHC gene TSS. Similarly, the 5’RACE analyses of the sense RNA targeting sense IIb RNA revealed the presence of multiple start sites with the major site corresponding at +1, but several minor sites clustered between -400 and -100 relative to the major TSS at +1. The close proximity of the antisense TSSs to those of the IIb MHC TSSs confirmed the idea that the two opposite-in-direction transcripts were initiated in the same area of the IIx-IIb intergenic region from a unique bidirectional
promoter or from two adjacent, head-to-head, opposing promoters. These findings strongly suggest that the IIx-IIb gene regulation is in a way similar to the cardiac β-α MHC genes whereby both sense and antisense transcription were detected in the intergenic region (20). The linear regression analyses showing that the xII NAT was strongly correlated with IIb pre-mRNA in the WMG when subjected to RE (R=+0.8 P<0.0001) (see Fig. 3 B) highlights the intriguing notion of common regulation between the two head to head situated promoters.

COMPARATIVE ANALYSES OF THE IIB MHC GENE PROMOTER SEQUENCE ACROSS SPECIES.

Based on the above results, the 400bp proximal promoter of the endogenous IIb MHC gene is transcriptionally active in two directions, and could be implicated in the regulation of not only the downstream IIb gene but also the upstream IIx gene as suggested by figures 3 and 4. Therefore, sequence analyses of the IIb promoter was undertaken in silico, using a comparative genomics approach, with a goal to gain further insight into its function such as identifying potential cis-elements / transcription factor binding sites (TFBS) that may be important for transcriptional regulation. With a larger availability of genomic sequences it is becoming clear that a high number of coding and structural sequences are conserved through many species. In addition, regions of noncoding DNA with a particularly high similarity among species are recognized as good candidates for functional significance, which is the basis of the evolving field of comparative genomics. Thus, we studied and compared the orthologous DNA sequences spanning the IIb MHC promoter from -1400 to +2800 bp relative to the IIb MHC TSS across four different species: Rat, Mouse, Human and Rhesus Monkey. The Rat sequence was used as the reference for the comparison.

Sequences were subjected to comparative analyses using the computational resources publicly available at http://mulan.dcode.org (36). Mulan is dynamically interconnected with
the multiTF utility (http://multitf.dcode.org) that identifies transcription factor binding sites (TFBS) that are shared among all the input species involved in the alignment. Thus, it is primarily designed to identify potential functional domains in a sequence predicated on sequence conservation. Analyses of the rat sequence against those sequences from 3 other mammalian species (mouse, rhesus monkey, and human; all extracted from the GenBank via Blast searches) identified two evolutionary conserved regions which shows >85% similarity across the 4 species (Figure 5). These two evolutionarily conserved regions (ECR) among all the four species are located in the area where the supposed bidirectional promoter is situated (figure 5). The first ECR was located between -294 bp and +47 bp relative to the sense IIb MHC TSS and included the first exon of the IIb MHC gene extending from +1 to +23. The second ECR was located at the 5’end of the IIb MHC gene, between +211 bp and 312 bp, and included the entire second exon extending from +224 to +275. Several muscle specific transcription factors (Myogenin, myoD, MEF-2, GATA) binding sites were found conserved across 4 species on ECR1, whereas both ECR1 and ECR2 displayed a variety of conserved binding sites for general and ubiquitous transcription factors such as C/EBP, SRF, CAAT, Sp1, Oct1, TBP, and these were located on both the sense DNA strand (+) and the complementary strand (-) (see figure 5B). While these observations are based on computational analyses of the sequence, the biological function of these associated binding sites and their role in skeletal MHC gene regulation remain to be determined in future studies. It is important to note that some of the identified TF binding sites on ECR1 by Mulan, matches well with what was determined before in studies focusing mainly on the mouse IIb promoter (4, 5, 43, 44), as well as the human IIb proximal promoter (26).
DISCUSSION

REGULATION OF THE SWITCH BETWEEN IIX AND IIB MHC ISOFORMS IN RESPONSE TO ACUTE RESISTANCE TRAINING

Skeletal muscle is the most abundant organ system in the body and, although considerable biochemical information has been already acquired concerning its structure and function, the complete knowledge of the molecular events underlying muscle plasticity has not been achieved yet. Therefore it is of great interest to identify and understand the molecular events that underlie the shift among myosin isoforms, which are responsible for the change in the fiber profile among targeted muscles.

In light of previous studies (17, 20, 38) and of our observations on the RNA expression by the use of strand specific RT-PCR approach, the goal of this study was to test the hypothesis that the switch from IIB to IIX fiber type in the WMG in response to RE is coordinated and regulated at the transcriptional level. Also we tested the hypothesis that this regulation involves a bidirectional promoter region, generating both sense and antisense RNA, thereby linking the regulation of the two genes. The RE paradigm used enabled us to observe a statistically significant modification in the MHC mRNA isoform composition in the WMG (Figure 1). These results are consistent with the findings that Caiozzo and colleagues reported (10), i.e., that RE induces an increase in IIX MHC expression with a concomitant decrease in IIB MHC in the whole WMG as well as in individual muscle fibers. The analysis of the primary transcription products of these two genes revealed the same pattern; i.e., the IIX pre-mRNA expression was up-regulated; whereas IIB pre-mRNA expression was down-regulated in the RE muscle (Figure 2). In addition, we also detected an antisense transcript overlapping with the IIX gene that was down regulated in RE muscles. The antisense transcript was detected corresponding to a long stretch on the IIX-IIb MHC gene locus. The RT-PCR results
and RNA mapping analyses suggest that the antisense RNA is a long transcript which originated from the IIb proximal promoter, spanned the IIx-IIb intergenic region, overlapped the entire IIx MHC gene and continued through the IIx promoter into the IIa-IIx intergenic region (Figure 4). In addition, analyses revealed a significant inverse relationship between the antisense transcript and the IIx pre-mRNA (Figure 3). These findings are consistent with the hypothesis that the antisense RNA is involved in the antithetical regulation of the IIx-IIb MHC genes, as seen for the cardiac α and β MHC genes and for the IIa-IIx genes (17, 20, 38). A closer examination of the inverse relationship between xII NAT vs IIx pre-mRNA shown in figure 3A reveals that when analyzing either the control group alone or the RE group alone, the correlation between the two variables is not significant. The lack of correlation in the normal control group may be due to the fact that when the xII NAT is expressed above a certain threshold, the IIx sense expression is blunted and maintained at a low level regardless of small fluctuation in xII NAT expression. On the opposite end, in the RE group, when xII NAT expression is reduced below a certain threshold, IIx sense gene expression is activated. However, in that state, any small fluctuation in the xII NAT is not correlated with those of IIx sense expression. In contrast, the negative correlation between xII NAT and IIx pre-mRNA becomes evident during dramatic shifts in IIx expression, such as in response to RE.

In the IIb-IIx switching model involving the white region of the MG, as depicted in figure 6, the xII NAT appears to act like an inhibitory cis-factor in the transcription process of the IIx gene. Non-coding antisense transcript expression “in cis” seems able to obstruct the sense transcription thereby silencing the overlapping gene; but how this happens precisely is still unresolved, but it likely involves epigenetic processes such as DNA methylation and chromatin remodeling. Recently, several RNA-dependent epigenetic processes have been described and are involved in transcription regulation. For example in yeast, antisense
transcripts mediate the PHO84 gene silencing via an epigenetic process involving the recruitment of histone deacetylases at the gene promoter, thus promoting a repressed state (12). In another instance, a chromatin remodeling factor, Isw2, has been implicated in repressing antisense transcription that may occur at the 3’ end of the genes in the yeast genome (52). Isw2 provides a regulatory mechanism to control gene expression by occluding the transcription start site or regulatory sequences through nucleosome repositioning along the DNA (52). Furthermore, antisense non-coding RNA has been implicated with chromatin silencing in the X-inactivation and in imprinting, and in the hypermethylation and silencing of CpG islands in promoters of target genes (see (47) for review).

It is intriguing to speculate that similar processes of noncoding-RNA dependent gene silencing are involved in MHC gene switching in muscle fibers, however, future research is needed to identify these processes.

STUDY OF THE LOCUS OF THE IIX-IIB INTERGENIC REGION

Our data show that the sense and the antisense transcripts share the same proximal promoter region (Figure 4). Through RNA mapping using a one step-RT PCR approach and 5’RACE analyses, we could circumscribe the area of transcriptional origin and limit it to a region of 403-271 bp relative to the major IIb TSS located at +1. The transcription start sites analysis carried out with a 5’RACE approach on the rat sequence indicated a cluster of multiple start sites for the antisense RNA in the area situated between -270 and -413 bp upstream of the IIb MHC TSS. This finding is not surprising and it is in agreement with the cardiac antisense β RNA in which at least two TSSs were identified (17) and the skeletal muscle antisense IIA RNA (aII NAT), whereby more than one TSS was found (38). In addition, mapping analyses and 5’RACE targeting the sense IIb RNA revealed that IIb transcription is carried out with the major start site at +1. However, several minor sites were also found in the IIb promoter,
which result in sense RNA that merges with the IIb pre-mRNA and is subjected to similar splicing as suggested by the PCR results using primer 10 (figure 4). The observation of the presence of multiple TSS for IIb is not surprising and confirms the finding by Dennehey et al., for the mouse gene showing that the IIb MHC mRNA is expressed in several isoforms that are different in their 5' untranslated region (14). Some of these RNA start as far as -1kb from the IIb TSS (+1) (14). The presence of a bidirectional promoters that co-regulates both the sense and antisense RNA transcription is very interesting and is common in human genes. Recently, it has been reported that over 10% of human genes are divergent, i.e., controlled by bidirectional promoters (1, 30, 46). Such promoters are defined as intergenic, less than 1kb long, and flanked by a transcription start site (TSS) on both of the DNA strands (32). What makes the bidirectional proximal IIb gene promoter even more interesting in the present study is that its activity is in an inverse relation with the activity of the upstream IIx sense gene (figure 3A and figure 6). That is, when the bidirectional promoter is active, both the IIb sense and the antisense IIx RNA are transcribed, whereas the IIx MHC gene is shut down. In contrast, when the bidirectional promoter undergoes inhibition, both the IIb gene and the antisense transcript are down-regulated, but the IIx gene located upstream is activated. There is a simultaneous ying-yang relationship, facilitated by cross-talk through more than 37 kb of genomic DNA. Therefore, the intergenic region that makes up the proximal bidirectional promoter may act not only as a physical linkage between the two MHC genes (IIx and IIb), but it seems to be essential also for their coordinated regulation. Therefore, the IIx and IIb MHC gene are physically and functionally linked through the intergenic bidirectional promoter region as suggested by the model presented in figure 6. In particular, it appears that in order for the IIx MHC gene to be regulated, a feedback from the intergenic antisense RNA is needed, especially in states whereby shifts in IIx-IIb expression occurs in response to a change is physiological stimuli.
COMPARATIVE SEQUENCE ANALYSIS

Genome sequence comparisons are typically used to identify noncoding genomic DNA regions that have been evolutionary conserved across species, presumably to maintain some critical biological function. Such conserved regions often correspond to transcriptional factor binding sites and regulatory regions (33, 35, 48). Thus, in this study, a comparative genomic approach was applied to provide insights on intergenic IIx-IIb promoter sequence conservation among species. MULAN, a computational tool available at http://mulan.dcode.org which is able to generate multiple sequence alignments, was used to identify conserved regulatory elements by comparing genomic sequences between related species (36). Four orthologous sequences of 4 kb length, spanning from −1.4 to +2.8 relative to the IIb TSS (+1), were compared. The rat sequence was analyzed against mouse, rhesus monkey and human DNA sequences. The sequence conservation through evolution presumes a regulatory function of that region. The comparative analysis showed in fact that, between -1.4 kb and +2.8 kb from the IIb MHC TSS, two regions were highly conserved (85% of similarity) (Figure 5). The first one (ECR1) was shown to be across the intergenic region and into the beginning of the IIb MHC gene, located from -294 bp to+47 bp (Figure 5). The second one (ECR2) was found within the 5’ end of the IIb MHC gene from +211 to+312 bp relative to IIb TSS (+1). Several muscle specific transcription factor-binding sites (TFBS) were found on ECR1 upper and lower strands; whereas, both ECR1 and ECR2 exhibited several general TFBS. The ECR1 analysis identified several potential cis regulatory elements such as CAAT box, C/EBP, SRF and the muscle specific TFBS for myogenin and MEF-2 (Figure 5). These conserved regulatory elements were found on the sense strand as well as on the reverse strand. Since ECR1 and ECR2 are highly conserved, their properties may be important promoter regions controlling both the antisense as well as the sense transcription.
The location of the TSS, the presence of conserved cis regulatory elements driving transcription on both the DNA strands, and also the position of the very high conserved region across four species provides evidence that the sequence considered (ECR1) is highly probable to function as the common promoter region for the sense IIb and the xII NAT.

It is worth to mention that previous studies on the mouse IIb promoter have identified some important cis regulatory regions that coincide with some of the ones conserved in ECR1 in the proximal promoter (see figure 5). For example, Takeda et al have identified three AT-rich regulatory regions and a proximal E box (43, 44), and later it was determined that these AT-rich sites interact with MEF-2 and Oct1 (27). Furthermore, the functional role of MEF2, MRF, and SRF transcting factors in the regulation of the IIb proximal promoter in muscle fibers was highlighted in studies by Allen et al (4, 5). Konig et al (26) have followed a comparative genomic approach to identify conserved domains in the proximal promoter of mouse and human IIb. However, these previous studies were mostly concerned with the regulation of the IIb promoter activity in the sense direction whereas in this current study we unraveled bidirectional activity and thus we are more concerned in finding that some of these factors can operate for both the sense and antisense transcription.

RELEVANCE TO IIx-IIb CO-EXPRESSION IN SINGLE FIBERS.

The IIx and IIb MHC genes are closely linked on the skeletal MHC gene locus, and in the white MG are regulated in opposite direction as the muscle undergoes phenotype shift in response to training. Both isoforms are expressed in fast skeletal muscle fibers (9), and therefore may respond to the same transacting factors. In fact, The similarity of TF binding sites in the proximal promoter among the fast MHC genes, in particular the MEF2 –AT rich sites was noted in several reports (5, 26, 43). Being part of a gene cluster, these genes may be regulated via a common enhancer (26), however, the antisense xII NAT may be important
when opposite regulation of the two genes is required. At present, it is not clear if within the same locus both IIx and IIb are being transcribed at different levels, such that in control white MG, 70-80% of the total MHC pool is IIb whereas the rest is IIx. On the other hand, because the muscle fibers are multinucleated, one might consider that the mosaic nature of MHC isoform expression within a single fiber is because individual myonuclei may be expressing different components of the sense and antisense regulatory circuit differentially at any given moment. On the other hand, because the muscle fibers are multinucleated, one might consider that the mosaic nature of MHC isoform expression within a single fiber is due to individual myonuclei expressing different components of the sense and antisense regulatory circuit differentially at any given moment. Another point to consider is that the dual expression of these two isoforms in many of the individual IIx/IIb fibers is not an all or none process. That is the antisense model does not operate in a fashion that IIb is fully repressed while the IIx is optimally expressed. Clearly the regulation is more complex as our data indicate.

RELEVANCE TO THE DIFFERENTIAL IIb MHC GENE EXPRESSION BETWEEN RODENT AND HUMAN

The results of this study show that xII NAT and the pre-mRNA of the IIb gene are generated by a unique bidirectional promoter located in the intergenic region flanking the 5’ end of the IIb gene. The IIx-IIb promoter region is highly conserved across mammalian species, even in species phylogenetically distant from each other such as rat and human. In view of the fact that the IIb MHC gene is rarely expressed in human skeletal muscle (24), it is intriguing to find out why the human IIb gene regulation is different from that of the highly expressed rodent IIb, considering that major regulatory components of the IIb promoter are conserved between the two species. Preliminary studies of the human IIb promoter in a transient reporter assay following direct gene transfer in skeletal muscle show that the human IIb
promoter is active in rodent fast muscle and exhibited muscle type specificity as the mouse IIb was shown previously (5). Thus, it would be useful and intriguing to ascertain the factor/s that really promote the transcription of the IIb MHC gene in rodent vs human muscle fibers.

SUMMARY
Collectively, this study provides significant insight on the dynamic regulation of the fast IIx and IIb skeletal muscle MHC genes. The expression of the two tandem genes, IIx and IIb, in the WMG is well coordinated to allow rapid shifts between the two isoforms, when needed in response to resistance loading. This regulation occurs at the transcriptional level and is tightly coordinated by a cis-acting natural antisense RNA (xII NAT). Transcription mapping of the IIx-IIb intergenic region revealed the generation of sense IIb and xII NATs from a single promoter region. This bidirectional promoter is highly conserved among species and contains several regulatory elements that may be implicated in its regulation. These results suggest that the IIx and IIb genes are physically and functionally linked via the bidirectional promoter generating antisense RNA which appears to be essential for the coordinated regulation of the skeletal muscle MHC genes during dynamic phenotype shifts. In order for the IIx MHC gene to be regulated, a feedback mechanism from the intergenic xII NAT is needed.
ACKNOWLEDGEMENTS

The authors would like to thank Li Ying Zhang, Phuc Tran, Alvin Yu, Bryce Buchowicz, Daniel Jimenez, Sandy Liu, Nkiruka Ojukwu, Jasleen Saini and Tiffany Yu for excellent technical assistance.

Grants.

Supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases Grant AR-30346 to K. M. Baldwin.
REFERENCES


FIGURE 1
MHC mRNA distribution profile as determined by competitive RT-PCR. Also shown are the PCR products separated on the gel and stained with ethidium bromide. MHC is the band amplified from the cDNA, the control is an external synthetic control added to the cDNA and amplified using the same MHC primers but generate a smaller size product (see ref (3, 15) for details). * P<0.05 RE vs. CON. RE: resistance exercise training; CON: normal control.
FIGURE 2
RNA expression for IIx and IIb MHC in CON and RE White MG as determined by strand-specific One-Step RT-PCR. A, IIx MHC pre-mRNA; B, IIx mRNA; C, antisense IIx RNA (xII NAT); D, IIb MHC pre-mRNA and E, IIb mRNA. Representative gels are shown in F for each of the RNA species. Antisense IIb RNA was below detection level; therefore, it is not shown. Sequence of primers used for these analyses is reported in supplementary table S2. *: P<0.05 RE vs. CON.
FIGURE 3
Linear relationship between antisense IIx RNA vs. IIx pre-mRNA (A) and vs. IIb pre-mRNA (B) in the white MG under CON or RE condition. Regression lines were generated by linear regression using Graphpad Prism software. R is pearson correlation coefficient.
FIGURE 4
Mapping RNA transcription across the IIa-IIx-IIb MHC gene locus. Gray arrows represent RNA detection using strand specific RT-PCR. Open boxes on the IIx-IIb intergenic region (IG) represent genomic repeats, whereas the hatched box represents a yet unknown sequence (N=661). Below is a high resolution mapping of the IIb proximal promoter which covers 1kb from the end of the IIx-IIb IG, and 1kb from within the IIb gene at its 5’ end. Bars 1 thru 13 are primer pairs used for mapping RNA expression via 1-step RT-PCR. Both sense (S) and antisense (AS) RNA are shown for each primer pairs for both control (CON) and trained (RE) WMG muscles, as well as genomic DNA amplification products. For each reaction, 100 ng of total RNA were used and PCR was run for 30 cycles. Note that for primers 2 thru 9, the gels show a light band corresponding to amplification of sense RNA, the light band is barely visible in the high contrasted image. The amplified band is however better visible when larger amounts of RNA were used (300ng/reaction, data not shown).

a: Primer 10 resulted in an intense band that is less than the predicted size. Sequencing of this lower molecular size band revealed that this is a spliced RNA, in which intron 1 of the IIb gene is spliced out. The shown band was 97 bp instead of expected 297 (supplementary table S3), and is down regulated with RE. See table S3 for primer sequence. For genomic DNA amplification, 20ng of rat genomic DNA was amplified for 27 cycles using Biolase DNA polymerase (Bioline).
FIGURE 5
Multispecies comparison of the IIX-IIb MHC promoter region. A) Graphic conservation profile obtained from mulan analyses (36) in the alignment of the 4 sequences using 85% similarity across at least 100 bp.

ECRs are displayed as dark blocks on top of each species conservation layer and they are linked to the underlying alignments in B. B) aligned sequence with identified conserved TFBS using mulan-MultiF function. (+) TFBS is on upper strand, (-) TFBS is on lower strand. * is 100% conservation of the corresponding nucleotide. MEF2: myocyte enhancer factor 2; CEBP: CAAT enhancer binding protein; SRF: serum response factor; SEF1: suppressor of essential function; PU1: is a member of the Ets family of transcription factors; ATF: activating transcription factor 1. Also shown in B are previously identified AT-rich regions on the mouse IIb promoter AT-1, AT-2, and AT-3 as designated by Takeda et al., (43) SRF is CArG box-1 in Allen et al (4, 5). TBP is also known as the TATA box, and the myogenin/MyoD sites were referred to as E-box (4)
The +2.8 fragment of the IIb gene contains Exon intron 1, exon 2, intron 2, and a part of exon 3. The ATG protein translation start site resides on the third exon for all these orthologous IIb genes.
FIGURE 6

Dynamic regulation of the IIx-IIb gene in the WMG muscle in response to acute RE. In control muscle, the bidirectional promoter (Bi-P) is active (++), IIb expression is high, whereas IIx expression is low, apparently inhibited by the antisense transcription. In trained muscle, conditions become unfavorable for the bidirectional promoter, its activity is inhibited (--), whereas the IIx gene promoter is simultaneously activated (++) . Gene spacing is not to scale.
Figure 1

![Figure 1 Diagram](image-url)
Figure 2

A

ILx pre-mRNA
AU/ng total RNA

CON RE

B

ILx mRNA
AU/ng total RNA

CON RE

C

xII NAT RNA
AU/ng total RNA

CON RE

D

ILb pre-mRNA
AU/ng total RNA

CON RE

E

ILb mRNA
AU/ng total RNA

CON RE

F

ILx pre-mRNA

ILx mRNA

xII NAT RNA

ILb pre-mRNA

ILb mRNA
Figure 4

[Diagram showing genetic structures and RNA detection patterns]

- IIa-IIx IG (~2.7kb)
- IIx MHC (~23kb)
- II-IG (~14kb)
- IIb MHC (~23kb)

S RNA
AS RNA
GDNA

CON - RE detection

Primer # 1 2 3 4 5 6 7 8 9 10a 11 12 13

+1 -1kb +1kb
Figure 6