Transcriptional Regulation of the Human Nonmuscle Myosin II Heavy Chain-A Gene

IDENTIFICATION OF THREE CLUSTERED CIS-ELEMENTS IN INTRON-1 WHICH MODULATE TRANSCRIPTION IN A CELL TYPE- AND DIFFERENTIATION STATE-DEPENDENT MANNER*

(Received for publication, November 26, 1997, and in revised form, January 16, 1998)

Nirat Beohar‡ and Sachio Kawamoto§

From the Laboratory of Molecular Cardiology, NHLBI, National Institutes of Health, Bethesda, Maryland 20892

Myosin is a family of mechanochemical proteins that contain a conserved ~80-kDa motor domain which can bind to actin, hydrolyze ATP, and translocate along actin filaments (1, 2).

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EBI Data Bank with accession number(s) AF033834.

‡ Present address: The George Washington University, Division of Cardiology, Washington, D.C.

§ To whom correspondence should be addressed: National Institutes of Health, Bldg. 10, Rm. SN202, 10 Center Dr., MSC 1762, Bethesda, MD 20892-1762. Tel.: 301-496-1912; Fax: 301-402-1542.

§§ To whom correspondence should be addressed: National Institutes of Health, Bldg. 10, Rm. SN202, 10 Center Dr., MSC 1762, Bethesda, MD 20892-1762. Tel.: 301-496-1912; Fax: 301-402-1542.

The myosin family now consists of 14 classes. Conventional myosin (myosin II in the new classification) consists of a pair of heavy chains (~200 kDa) and two pairs of light chains (15–20 kDa) and can form filaments through an α-helical coiled coil rod-like region. While nonmuscle myosin II is present in all eukaryotic cells, in higher organisms, different types of cells contain different isofoms of myosin II. In vertebrates, there are over 10 different isofoms of myosin II that are divided into two subgroups, i.e. sarcomeric (skeletal and cardiac muscles) and nonsarcomeric (smooth muscle and nonmuscle) myosins. The contractile property of sarcomeric and smooth muscle myosin II is a prominent feature and they serve as an integral part of the contractile apparatus in muscle cells. On the other hand, the exact function of nonmuscle myosin II is still under study, but numerous biochemical and cell biological studies, as well as genetic studies using Dictyostelium, have demonstrated that myosin II in nonmuscle cells are involved in diverse cellular motile processes including cytokinesis, capping of surface receptors, and cell shape changes (see Ref. 3 for review). Recent genetic studies using Drosophila and mouse systems have also demonstrated that nonmuscle myosin II plays a critical role in embryonic morphogenesis and development (4, 5).

Different isofoms of myosin II contain different myosin II heavy chains (MHCs) which are encoded by different genes. There are at least 8 genes for sarcomeric MHCs in vertebrates and expression of these genes is regulated developmentally, hormonally, and in a muscle fiber type-specific manner. This regulation occurs mainly at a transcriptional level and a number of muscle-specific enhancers and transcriptional factors have been identified (6–9). For nonmuscle MHCs, we and others demonstrated the existence of two genes, referred to as nonmuscle MHC-A and MHC-B genes, by cDNA cloning (10–15). In the case of nonmuscle MHC-B, an alternative pre-mRNA splicing mechanism is also utilized to generate additional nonmuscle MHC-B isoforms (15, 16). Human nonmuscle MHC-A and -B genes are located on chromosomes 22q11.2 and 17p13, respectively (12–14). The two nonmuscle MHC mRNAs are expressed in a variety of tissues, but the relative amounts of the two mRNAs vary among different tissues (10, 17). MHC-A is the dominant isofom in intestinal epithelium, spleen, and thymus, whereas MHC-B is dominant in brain and testis. Lung and kidney contain approximately equal amounts of each MHC mRNA. In contrast, both nonmuscle MHC-A and -B mRNAs are barely detectable in fully developed skeletal muscles where the sarcomeric MHC is dominant. Serum and other mitotic stimulants change the expres-
sion of these genes differently (14, 17). For example, serum stimulation up-regulates MHC-A gene expression, whereas it down-regulates MHC-B gene expression in fibroblasts. Accumulating evidence shows that the expression of the specific nonmuscle MHC isoforms is dependent on cell types and is linked to cell proliferation and differentiation, especially in neuronal and muscle cells. However, the regulatory mechanisms controlling the expression of nonmuscle MHC genes have not yet been elucidated.

We have recently cloned and characterized the promoter region of the human nonmuscle MHC-A gene (18). The structure of this region shows many features typical of a housekeeping gene. There is no TATA element and the GC content is high, with multiple GC boxes. Weir and Chen (19) have isolated genomic clones which encode the promoter regions of human and mouse nonmuscle MHC-B genes. These genes also lack a TATA element and the GC content is high. The finding that the nonmuscle MHC-A and -B genes belong to the housekeeping gene family, based on sequence and structural features of the promoter regions, is consistent with the previous reports that both genes are expressed in a wide variety of cell types and tissues. However, as noted, there are differences in expression of these genes among different cell types and tissues. Thus, we searched for the cis-regulatory elements that might be responsible for the regulation of nonmuscle MHC-A gene transcription.

We report here a cell type-dependent enhancer activity found in intron 1, located in the 23-kb downstream region from the transcriptional start sites of the human nonmuscle MHC-A gene. We identified three clustered protein-binding elements, one of which is recognized specifically by the Sp1 and Sp3 transcriptional factors.

EXPERIMENTAL PROCEDURES

Plasmid Construction—The core promoter luciferase reporter plasmid contains the core promoter of the human nonmuscle MHC-A gene which corresponds to the sequence between -1122 and +61 (where +1 is a major transcription start site) inserted at a HindIII site in a promoterless luciferase reporter vector pGL2 basic (Promega) and has been described (18). This plasmid was used as a host vector to make all constructs described below.

Constructs 5.9(+) (Fig. 1) and 2.2(+) (see Fig. 1) were constructed by inserting a 5.9-kb BglII fragment in intron 1 (see Fig. 1) into the core promoter luciferase plasmid at a BglI site, located upstream to the core promoter, in the same (+) and opposite (−) orientations. Constructs 0.9(+) (Fig. 1), 2.8(+) (see Fig. 1), and 2.2(−) (see Fig. 1) were constructed by deleting appropriate restriction fragments from constructs 5.9(+). Constructs 0.9(+) (Fig. 1) and 2.8(+) (see Fig. 1) were constructed by deleting appropriate restriction fragments from constructs 5.9(+). Constructs 0.9(+) (Fig. 1) and 2.8(+) (see Fig. 1) were constructed by deleting appropriate restriction fragments from constructs 5.9(+). Constructs 0.9(+) (Fig. 1) and 2.8(+) (see Fig. 1) were constructed by deleting appropriate restriction fragments from constructs 5.9(+).

RESULTS AND DISCUSSION

The Distal Downstream Region in Intron 1 Modulates Nonmuscle MHC-A Transcription in a Cell Type- and Differentiation State-Dependent Manner

We previously isolated genomic clones which encode the promoter and flanking region (~70 kb) of human nonmuscle MHC-A and characterized the core promoter and proximal regulatory regions (18). In an attempt to identify cell type-specific cis-regulatory elements, we now examined the region extending ~20 kb upstream and ~40 kb downstream from the transcriptional start site, which includes the 37-kb intron 1.

The genomic DNA clones were fragmented by restriction enzymes (BglII and BamHI) and introduced upstream to the MHC-A core promoter in the luciferase reporter constructs. Following transfection into NIH 3T3 fibroblasts, a 5.9-kb fragment in intron 1, that is located 21–27 kb downstream from the transcriptional start sites, was able to enhance transcrip-
Regulation of Nonmuscle Myosin Heavy Chain Gene Transcription

In the bottom panel promoter (BC2C12 myotubes. The relative luciferase activities normalized by muscle MHC-A gene. The distal downstream region located in intron 1 of the non-muscle MHC-A gene. The indicated fragment of the MHC-A gene, shown in the upper panel, is inserted 5’ to the MHC-A core promoter for the luciferase reporter construct in the same (+) or opposite (−) orientation. The various luciferase constructs were co-transfected with pSV-β-galactosidase (or pRSV-β-galactosidase) into NIH 3T3, BC2C12 myoblasts, and BC2C12 myotubes. The relative luciferase activities normalized by β-galactosidase activities are shown as bar graphs (mean ± S.E., n = 3–7) in the bottom panel. The luciferase activity due to the MHC-A core promoter (core prom.) in individual cell types is represented as 1.

2.8-kb fragment causes a 3–6-fold increase in luciferase activity in proliferating BC2C12 myoblasts and the repressor activity observed in differentiated BC2C12 myotubes are located in distinct regions within a 5.9-kb fragment. In this report, we now focus on identifying and characterizing the cis-acting elements and trans-acting factors responsible for high expression of the nonmuscle MHC-A gene in fibroblasts.

In an effort to localize the region responsible for the enhanced activity due to the 2.8-kb fragment, this fragment was progressively narrowed down using PCR generated overlapping fragments which were introduced into the luciferase reporter construct in both orientations. As shown in Fig. 2, the first set of analyses (N1-N9) demonstrates that three 450-bp fragments, N2, N3, and N7, possess enhancer activity in an NIH 3T3 background. The second set of reporter constructs, N2a-N2f and N7a-N7d, cover the region N2 plus N3 and N7, respectively. Analysis of these constructs defines a 100-bp fragment, N2d, with maximum enhancer activity (−20-fold) (Fig. 2). The nucleotide sequence of the N2d fragment is shown in Fig. 13. The activity of the N2d fragment still shows cell type dependence (Fig. 1) and is represented as 1, SVP, SV40 early promoter, SVPE, SV40 early promoter with enhancer.

2-fold decrease in transcription in BC2C12 myotubes, whereas this has no effect in NIH 3T3 and BC2C12 myoblasts. Thus, the enhancer activity observed in NIH 3T3 cells and proliferating myoblasts and the repressor activity observed in differentiated BC2C12 myotubes are located in distinct regions within a 5.9-kb fragment. In this report, we now focus on identifying and characterizing the cis-acting elements and trans-acting factors responsible for high expression of the nonmuscle MHC-A gene in fibroblasts.

In an effort to localize the region responsible for the enhanced activity due to the 2.8-kb fragment, this fragment was progressively narrowed down using PCR generated overlapping fragments which were introduced into the luciferase reporter construct in both orientations. As shown in Fig. 2, the first set of analyses (N1-N9) demonstrates that three 450-bp fragments, N2, N3, and N7, possess enhancer activity in an NIH 3T3 background. The second set of reporter constructs, N2a-N2f and N7a-N7d, cover the region N2 plus N3 and N7, respectively. Analysis of these constructs defines a 100-bp fragment, N2d, with maximum enhancer activity (−20-fold) (Fig. 2). The nucleotide sequence of the N2d fragment is shown in Fig. 13. The activity of the N2d fragment still shows cell type dependence (Fig. 1). Maximal activation due to the N2d fragment is seen in NIH 3T3 (−20-fold). BC2C12 myoblasts show intermediate activation (−7-fold). In contrast, transcription in...
of the MHC-A gene, see Fig. 1. R, the gel mobility shift assay using nuclear extracts from NIH 3T3 cells (Nuc. Ext.) and the 100-bp fragment N2d as a probe. Position of the two DNA-protein complexes (I and II) and free probe (F) are indicated. + and − represent presence and absence of probes and competitors, respectively. Concentrations of unlabeled competitor DNAs are shown as molar excess (× 10 and 100) relative to the labeled probe concentration. C, the gel mobility shift assay using the NIH 3T3 nuclear extracts and the 30-bp indicated subsegments as probes. N2d1 forms multiple complexes with similar mobilities, resulting in broad shifted bands (III, lane 2). N2d5 forms an apparent single complex (IV, lane 14). The competitors used in this assay were a × 10 molar excess of unlabeled probes.

To characterize these DNA-protein complexes further, we made use of a high resolution gel electrophoresis system with higher acrylamide gel concentrations. Using the N2d1 probe and NIH 3T3 nuclear extracts, as shown in Fig. 4, one major DNA-protein complex (a) is observed with a low amount of nuclear proteins (lane 1). With increasing amounts of nuclear proteins, three complexes b, c (appears as doublet c1 and c2), and d are detected (lanes 2–5). With the highest amount of nuclear proteins (lane 5), complex d becomes the dominant species, while complex a becomes a minor species. These results suggest that the N2d1 probe can recruit different proteins or different protein complexes through either a single or multiple protein-binding sites.

To examine if these complexes bind to a distinct region of the N2d1 probe, methylation interference assays were carried out for each of the complexes. Strong interference with protein binding due to methylation of G residues (and some A residues) is indicated by solid circles and weak interference by open circles in Fig. 5. These marked residues contact proteins and are required for protein-DNA complex formation. The complexes a and c1 + c2 bind to two distinct sequences within the 30-bp N2d1 probe. These binding sites are designated element A (GTGACCC) for complex a and element C (GGGAGGGGCC) for complex c1 + c2. Interestingly, complex d binds to both elements A and C.

We next evaluated how mutations in each of the elements affects formation of each of the N2d1-protein complexes (Fig. 6). The unlabeled wild-type N2d1 competes effectively with the labeled N2d1 probe for all complex formations, resulting in disappearance of all labeled N2d1-protein complexes, as expected (lanes 2 and 3). The mutant m1, which has mutations in element C, can compete with the N2d1 probe for formation of complexes a, b, and d, whereas it does not affect c1 and c2 complex formation (lanes 4 and 5), implying that complexes a, b, and d require element A, but complexes c1 and c2 do not. On
the other hand, mutant m2, which has mutations in element A, competes for formation of the c1, c2, and d complexes, but not for formation of a and b complexes (lanes 6 and 7), implying that element C is required for c1, c2, and d complexes, but not for a and b complexes. The mutant m3, which is mutated in both elements A and C, does not affect any complex formation, indicating that no other sequences besides elements A and C are required for complex formation. These results are consistent with the data obtained from methylation interference assays and establish that complex a requires only the A element whereas complexes c1 and c2 require only the C element and complex d requires both A and C elements. Moreover, the above data also indicate that b complex shares the same binding sequences with complex a, although the b complex was not analyzed by methylation interference assays.

Element C, GGGAGGGGCC, Interacts with Sp1 and Sp3—The sequences of the two elements A and C do not show complete identity with any known binding sites for transcription factors. However, the C element sequence (GGGAGGGGCC) resembles the consensus sequence for the Sp1-binding site (GGGGCGGGGC) (22). Therefore, the authentic Sp1-binding sequence as well as other known transcription factor binding sequences, which are rich in G residues, were tested for their effects on N2d1-protein complex formation in gel shift assays. As shown in Fig. 7, excess amounts of the unlabeled consensus Sp1-binding sequence is able to compete specifically for complexes c1, c2, and d, but not for complexes a and b using the N2d1 probe (lane 3). In contrast, these complexes are not affected by the presence of the same amounts of a mutated Sp1-binding sequence (Sp1m) (lane 4), demonstrating the specificity of competition with the Sp1-binding site. The binding sequences for AP2 (GCCCGCGG) and Egr (GGGGCGGGGC) also do not compete for the N2d1-protein complex formation with the N2d1 probe (lanes 5 and 6). Thus, only the Sp1-binding sequence is able to compete specifically for formation of complexes c1, c2, and d, consistent with the fact that formation of all three of these complexes requires element C. These observations prompted us to explore the possibility that Sp1 or Sp1-related protein(s) may interact with binding element C located in N2d1.

In addition to the Sp1 protein, three other proteins, Sp2, Sp3, and Sp4, which are structurally closely related to Sp1, have been shown to bind to the same DNA motif as Sp1 (23, 24). Therefore, antibodies specific for each of the four Sp family proteins were used to examine whether any Sp family proteins are components of complexes c1, c2, and d. As shown in Fig. 8A, in the presence of antibodies specific for Sp1, most of complex d and complex c1 change their mobilities and are retarded in migration (SS, supershift, lanes 1 and 7). This result suggests that anti-Sp1 antibodies recognize a component of complexes d and c1 and form higher order multiprotein-DNA complexes, resulting in slower migration in native gels. In the presence of antibodies specific for Sp3, part of complex d and complex c2...
change their mobilities and migrate to the top of the gel, as well as a band just below complex d (see small arrowheads alongside lanes 4, 8, and 9), suggesting that anti-Sp3 antibodies recognize a component of complexes d and c2 and result in slower migrations. Addition of both anti-Sp1 and anti-Sp3 antibodies causes supershifts of almost all of complexes c1, c2, and d (lane 9). In contrast, anti-Sp2 and anti-Sp4 antibodies have no effect (lanes 3 and 5). These results indicate that Sp1 and Sp3 (or Sp1-like and Sp3-like molecules), but not Sp2 and Sp4, are major components of complexes c1, c2, and d formed with probe N2d1, consistent with the fact that formation of all these complexes are competed with the consensus Sp1-binding sequence (see Fig. 7). Since formation of all these three complexes, c1 and c2, and d, has been shown to require the element C (GGGAGGGGCC), this is recognized by Sp1 and Sp3. This notion is further supported by data shown in Fig. 8B. When the mutant m2, which contains an intact element C, but a mutated element A in the context of N2d1, is used as a probe, m2 forms a major doublet complex (c1 and c2), which comigrates with complexes c1 and c2 detected by probe N2d1, and a minor complex m2c with faster mobility (lanes 1 and 6 in Fig. 8B). Anti-Sp1 antibodies supershift all of complex c1, but not complexes c2 and m2c, to the two slower migrating broad bands indicated by SS (lanes 2 and 7), demonstrating that anti-Sp1 antibodies recognize a component of complex c1. On the other hand, anti-Sp3 antibodies supershift both complexes c2 and m2c leaving the c1 complex at the same position, indicating that anti-Sp3 antibodies recognize a component of complexes c2 and m2c (lanes 4 and 8). These results also suggest that m2 can interact with either Sp1 or Sp3, but not both proteins per one DNA molecule. Addition of both anti-Sp1 and anti-Sp3 antibodies leads to supershifting of all complexes (c1, c2, and m2c) completely (SS, lane 9), indicating that no other factors in NIH 3T3 extracts can interact with the m2 probe. Anti-Sp2 and anti-Sp4 antibodies have no effect (lanes 3 and 5). Since complexes c1 and c2 bind to element C, the above data establish that element C, GGGAGGGGCC, is recognized specifically and exclusively by Sp1 and Sp3 transcription factors in NIH 3T3 nuclear extracts. For reference, the authentic Sp1-binding site was also tested in a gel mobility shift assay with NIH 3T3 nuclear extracts and the antibodies used in the above experiments. Results very similar to those with the m2 probe were obtained (data not shown).

As far as we are aware, sequences similar to element A (GTGAGCC) have not been reported previously. Therefore, this element may interact with a potentially novel transcription factor. Taken together, the 30-bp N2d1 region consists of two protein-binding elements: the more 5’-element C (GGGAGGGGCC) is recognized by Sp1 and Sp3 and the other element A (GTGAGCC), immediately preceded by element C, seems to be novel. Since complex d also contains Sp1 and Sp3 as its components, like complexes c1 and c2, but migrates slower than complexes c1 and c2, and formation of this complex requires an additional element A, complex d is most likely a multiprotein complex. Thus, the DNA fragment, which con-
contains both elements A and C, can form a multiprotein complex occupying both elements in addition to complexes in which either A or C is occupied.

Participation of E-box-binding Proteins—Next, we characterized the N2d5-protein complex detected using the 30-bp N2d5 probe and NIH 3T3 nuclear proteins (complex IV in Fig. 3). In a high resolution gel electrophoresis system, the N2d5-protein complexes are resolved into a major complex f1 and a minor complex f2 with a slightly faster mobility, as well as another minor complex f3 (lane 1 in Fig. 9). The unlabeled wild-type N2d5 competes with the labeled N2d5 probe for formation of all three complexes, showing that all three complexes are specific (lanes 2 and 3). Methylation interference analysis of complexes f1 and f2 shown in Fig. 10 reveals that the protein-binding sequence is GTGTCAGGTA, which is designated element F. Competition experiments using mutant m4, which is mutated at element F in the context of N2d5, demonstrate that formation of none of the three complexes is inhibited by the presence of m4 (lanes 4 and 5). This indicates that complexes f1-f3 share the same binding sequence. Element F is located 50 bp 3' to element A in the N2d region.

Of note, element F, GTGTCAGGTG, includes an E-box motif, CANNTG (where N is A, T, G, or C). The E-box is known to be recognized by transcription factors that contain a basic-helix-loop-helix motif or a basic-helix-loop-helix-leucine zipper motif. Therefore, we examined the possible involvement of a number of known E-box-binding proteins in the N2d5-protein complexes, by competition experiments, using authentic binding sequences for specific E-box-binding proteins and antibody supershift experiments in gel shift assays. Fig. 11A shows that the authentic Myc-Max and USF (upstream stimulatory factor) binding sequence, both of which contain the sequence CACGTG (25–27), but not the mutated sequence (Myc-Maxm and USFm), compete specifically for f2 complex formation (lanes 2–7). On the other hand, the MEF-1 binding sequence (28) has no effect, even though this binding sequence contains the same

![Fig. 8A](https://example.com/image8a.png)

**Fig. 8A.** Sp1 and Sp3 are components of the N2d1-protein complexes. Antibodies specific to transcriptional factors Sp1-Sp4 are used in gel mobility shift assays, performed using the N2d1 probe and NIH 3T3 nuclear extracts. a-d indicate DNA-protein complexes formed in the absence (−) of specific antibodies (Ab.). The upper arrowhead labeled SS indicates the supershifted complexes formed resulting from binding of anti-Sp1 and anti-Sp3 antibodies (lanes 2, 4, and 7–9). The lower arrowheads labeled SS as well as the small arrowheads alongside lanes 4, 8, and 9 indicate supershifted bands due to binding of anti-Sp3 antibodies. B. Sp1 and Sp3 can bind element C, GGGAGGGGCC, in N2d1. Mutant m2, in which element A is mutated, but element C is intact in the context of N2d1 (see “Experimental Procedures” for sequence), is used as a probe with NIH 3T3 nuclear extracts in gel mobility shift assays. DNA-protein complexes c1, c2, and m2c formed in the absence of antibodies (Ab. −) are indicated. A doublet complex c1 and c2 comigrates with complexes c1 and c2 formed by probe N2d1 (see Fig. 4). Indicated antibodies (Ab.) are added to the reaction mixtures. The upper arrowheads labeled SS indicate supershifted complexes formed resulting from binding of anti-Sp1 and anti-Sp3 antibodies (lanes 2, 4, and 7–9). The lower arrowheads labeled SS indicate supershifted complexes due to binding of anti-Sp3 antibodies (lanes 2, 7, and 9).

![Fig. 9](https://example.com/image9.png)

**Fig. 9.** Gel shift assay demonstrating specific complex formation with N2d5. The DNA-protein complexes f1-f3 formed by the 30-bp N2d5 probe and NIH 3T3 nuclear extracts and free probe are indicated. The wild-type N2d5 and mutant m4 in which element F (see Fig. 10) is mutated (see “Experimental Procedures” for sequence) are used as competitors (Comp.). The concentrations of unlabelled competitors are shown as molar excess to the labeled probe.
core hexamer sequence CAGGTG, as element F. Fig. 11B demonstrates that antibodies against USF1 and USF2, but not against c-Myc or Max, cause supershifting of the f2 complex to the top of the gel (SS, lane 8). The other antibodies against the ubiquitously expressed E protein family (E47, E12, and E2A) do not affect mobility of any complexes. Together with the results shown in Fig. 11, A and B, these results demonstrate that USF1 and/or USF2 are components of complex f2. Thus, the E-box sequence located in N2d5 is partially recognized by USF1 and/or USF2. The major factor(s) which form the f1 complex, however, has not yet been identified.

Cooperative Effects of Cis-elements on Transcriptional Activation in Fibroblasts

The in vitro DNA-protein binding analyses described above demonstrate the existence of three binding elements within the 100-bp N2d region. Finally, we evaluated the contribution of these binding elements for activation of nonmuscle MHC-A gene transcription. To this end, luciferase reporter constructs, which include mutations at three elements in various combinations, were created in the context of the 120-bp N2d’ region (see Fig. 13 for the sequence of the N2d’) and were transiently transfected into NIH 3T3 cells. All mutations created in the luciferase constructs contain the identical nucleotide substitutions to those in mutants m1-m4, which have been shown to abolish protein binding in gel shift assays. The data from luciferase assays due to each construct are summarized in Fig. 12. Wild-type N2d’ causes an ~30-fold increase in transcriptional activity compared with that due to the core promoter alone (line 1 in Fig. 12). Mutation of all three elements abolish enhancer activity almost completely (line 8), suggesting that no other DNA sequences besides the three elements within the N2d region contribute to transcriptional enhancement. When two other elements are mutated, the individual elements C, A, and F alone show transcriptional enhancer activity 4.9-, 35.8-, and 3.3-fold (in an average of two different orientations), respectively (lines 7, 6, and 4). Comparing lines 1 versus 5, 2 versus 6, 3 versus 7, and 4 versus 8, the effect of element F is consistent with being additive to the effects due to either element A, C, or both. However, the combined effects of elements A and C are somewhat complicated. Mutations in element A without mutations in element C causes a decrease in transcript-
tion activity to about 65% of that due to the constructs in which both A and C are intact, regardless of the presence or absence of mutations in element F, indicating that element A contributes only to a 1.5-fold increase of activity (compare lines 1 versus 3 and 5 versus 7). On the other hand, surprisingly, mutation of element C (but not element A, with or without mutation of F) results in a 3–5-fold increase in activity compared with the constructs in which both A and C are intact (compare lines 1 versus 2 and 5 versus 6). This implies that element C functions as a repressor in the presence of an intact element A. As pointed out earlier, however, element C itself is also able to activate transcription in the absence of an intact element A. Thus, the effects due to elements A and C are not additive, but rather cooperative. In addition, differences in the degree of mutational effects seen in the two different orientations (normal and reverse) of the N2d5 region on transcriptional activity. A number of mechanisms can be imagined whereby Sp1 and/or Sp3 binding to element C could activate or inhibit transcription. Sp1 is one of the best characterized transcriptional activators (22, 29). On the other hand, Sp3 has recently been described as a protein which can bind to GC or GT boxes, similar to Sp1. In contrast to Sp1, however, the functional effects of Sp3 are controversial (30–33). Co-transfection studies of Sp3 expression vectors with promoter-reporter constructs showed that Sp3 functions as a repressor for some promoters while it functions as a activator for other promoters in different cellular backgrounds. Thus, the effects of Sp3 varies depending on the system used. Moreover, Majello et al. (34) reported that

### Transcriptional Activity (fold increase, core promoter = 1)

| Elements | Normal | Reverse | Average |
|----------|--------|---------|---------|
| 1. C     | 31.6±4.6 | 28.7±5.8 | 30.1 (100%) |
| 2. CA    | 71.6±14.0 | 94.5±20.0 | 83.1 (276%) |
| 3. A     | 24.4±4.5 | 16.4±3.0 | 20.4 (67%) |
| 4. C     | 5.3±0.77 | 1.3±0.09 | 3.3 (11%) |
| 5. CA    | 6.7±0.95 | 8.3±1.6 | 7.5 [100%] |
| 6. A     | 21.0±3.6 | 50.5±10.9 | 35.8 [477%] |
| 7. C     | 5.3±0.63 | 4.4±0.49 | 4.9 [65%] |
| 8. CA    | 2.0±0.27 | 0.79±0.06 | 1.4 [18%] |

○: wild type
×: mutation

---

**Fig. 11.** Participation of E-box-binding proteins in the N2d5-protein complex. A, competition of the consensus sequences for E-box binding transcription factors for N2d5-protein complexes. The consensus binding sequences for E-box binding proteins, Myc-Max, USF, and MEF-1, as well as their mutant sequences (Myc-Maxm, USFm, and MEF-1m), are used as competitors for N2d5-protein complex formation with NIH 3T3 nuclear extracts in gel mobility shift assays. DNA-protein complexes f1-f3 are indicated. The concentrations of unlabeled competitors are shown as molar excess to the probe concentration. B, USF or USF-related protein is a component of complex f2. Antibodies specific to E-box binding transcriptional factors indicated (Ab) are used in gel mobility shift assays performed using the N2d5 probe and NIH 3T3 nuclear extracts. Complexes f1-f3 formed in the absence of antibodies (Ab–) are indicated. An arrowhead with SS represents supershifted complexes formed due to binding of antibodies specific to USF1 and USF2 (lane 8).

**Fig. 12.** The effect of mutating cis-acting elements located in the N2d region on transcriptional activity. Three cis-acting elements, A, C, and F, were mutated in various combinations in the context of the 120-bp N2d′ region. The wild-type and mutated DNA fragments were inserted 5′ to the MHC-A core promoter for the luciferase reporter constructs in two different orientations (Normal, Reverse). The various luciferase constructs were co-transfected with pSV-β-galactosidase into NIH 3T3 cells. The relative luciferase activities normalized by β-galactosidase activity are represented as fold increase (mean ± S.E., n = 7), compared with the luciferase activity due to the MHC-A core promoter (1). a, % activity comparing lines 1–4; b, % activity comparing lines 5–8.
the Sp3 protein contains independent modular repressor and activator domains. The predominant effect of Sp3 seems dependent on the context of the Sp3 DNA-binding sites and on the nature of the particular cellular backgrounds. Kennett et al. (35) also reported that Sp3 mRNA encodes multiple polypeptides having a different initiating methionine codon within the same reading frame. These multiple Sp3 proteins differ in their capacity to activate or repress transcription. Indeed, we constantly observed that the m2 probe, as well as the authentic Sp3 probe (data not shown), forms two complexes, c2 and m2c, which are recognized by antibodies specific to Sp3 (Fig. 5B). These complexes may be due to the different Sp3 isoforms which result from different initiation sites for translation. Of these isoforms, the possible involvement of Sp3 proteins in transcriptional regulation via the N2d region may explain, in part, why the effects of element C differ depending on the DNA context.

Although we demonstrated that both Sp1 and Sp3 can interact with element C in vitro, we do not know whether Sp1 and Sp3 are randomly binding to a target DNA element, depending on their relative concentrations, or if there are mechanisms for discriminating among Sp1 and Sp3 proteins for DNA binding in intact cells. The DNA-binding domains of Sp1 and Sp3 proteins share very similar primary structures and the core target DNA sequences tested so far show similar affinities to two factors in vitro (24). However, the binding affinity of each protein may be modified by the context of the DNA, especially when it includes target sequences for other factors. Of particular note is that Sp-binding element C is immediately followed by element A in the N2d enhancer region of nonmuscle MHC-A. We were able to demonstrate that the N2d1 probe can form a multicomponent-DNA complex, d, which occupies both elements A and C (Figs. 4 and 5). Based on the results of the antibody supershift experiments, both Sp1 and Sp3 can be components of complex d (Fig. 8A). We have not yet identified the factor(s) which binds to element A. It is known that Sp1 physically interacts and cooperates with several transcription factors such as YY1, p53, GATA-1, E2F, and Egr (36–40). Element A is not a typical binding site for these proteins nor for the other transcription factors. It will be necessary to identify the element A-binding factor to understand the activation mechanism of nonmuscle MHC-A gene transcription via the N2d enhancer region. It is also worthwhile to point out that the nonmuscle MHC-A core promoter contains putative Sp-binding sites (18). Synergistic activation through multiple Sp1 proteins described in other genes (41) may also contribute to full activation of the nonmuscle MHC-A promoter.

In summary (see Fig. 13), we identified three clustered cis-acting elements within the 100-bp region that are located 23 kb downstream from the transcriptional start sites in the first intron. The most 5’ element C (GGGAGGGGCC) is recognized exclusively by Sp1 and Sp3 and is immediately followed by a novel element A (GTGACCC). The third element F (GTGTACAGGTG) is located 50 bp 3’ to element A and contains an E-box. This element can be recognized partially by USF, but the major factor(s) binding to this element has not been identified. Transfection studies with luciferase reporter constructs which include mutations in all three elements in various combinations demonstrate that the A and C binding factors cooperatively activate transcriptional activity. The F binding factor shows an additive effect. These factors appear to be responsible for high expression of nonmuscle MHC-A gene in fibroblasts. A more detailed understanding awaits the identification and characterization of the factors which bind to elements A and F. Nevertheless, characterization of the Sp-binding site in the strong enhancer region and its interaction with Sp1 and Sp3 helps us to define the molecular interplay among multiple cis-acting elements and trans-acting factors in the regulation of the nonmuscle MHC-A gene transcription.

Acknowledgments—We are grateful to Dr. Robert S. Adelstein (NHLBI) for continued encouragement, helpful discussions, and critical reading of the manuscript. We also acknowledge the useful contribution of Arlen Pyenson and Justin Bekelman (Princeton University), and the excellent editorial assistance of Catherine S. Magruder.

REFERENCES
1. Moseker, M. S., and Cheney, R. E. (1995) Annu. Rev. Cell Dev. Biol. 11, 633–675
2. Sellers, J. R., and Goodson, H. V. (1995) Protein Profile 2, 1233–1423
3. Spudich, J. A. (1989) Cell Regul. 1, 1–11
4. Young, P. E., Richman, A. M., Ketcham, A. S., and Kiehart, D. P. (1993) Genes Dev. 7, 29–41
5. Tullio, A. N., Accili, D., Ferrans, V. J., Yu, Z.-X., Takeda, K., Grinberg, A., Westphal, H., Preston, Y. A., and Adelstein, R. S. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 12407–12412
6. Weintraub, H. (1993) Cell 75, 1241–1244
7. Olson, E. N. (1993) Circ. Res. 72, 6–10
8. Yu, Y.-T., Broichard, R. E., Smoot, L. B., Lee, Y., Mahdavi, V., and Nadasdy, V. (1992) Genes Dev. 6, 1783–1798
9. Robbins, J., Palermo, J., and Rindt, H. (1995) Ann. N. Y. Acad. Sci. 752, 492–505
10. Katsuragawa, Y., Yanagisawa, M., Inoue, A., and Masaki, T. (1989) Eur. J. Biochem. 184, 611–616
11. Shohet, R. V., Conti, M. A., Kawamoto, S., Preston, Y. A., Brill, D. A., and Adelstein, R. S. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 7726–7730
12. Saenz, C. G., Myers, J. C., Shows, T. B., and Leinwand, L. A. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 1364–1368
13. Simons, M., Wang, M., McBride, O. W., Kawamoto, S., Yamakawa, K., Gdula, D., Adelstein, R. S., and Weir, L. (1991) Circ. Res. 69, 530–539
14. Tosthaker, L. E., Gonzalez, D. A., Tung, N., Lemons, R. S., Le Beau, M. M., Arnaout, M. A., Clayton, L. K., and Tenen, D. G. (1991) Blood 78, 1826–1833
15. Takahashi, M., Kawamoto, S., and Adelstein, R. S. (1992) J. Biol. Chem. 267, 17864–17871
16. Kawamoto, S. (1996) J. Biol. Chem. 271, 17613–17616
17. Kawamoto, S., and Adelstein, R. S. (1991). J. Cell Biol. 112, 915–924
18. Kawamoto, S. (1984) J. Biol. Chem. 259, 15101–15110
19. Weir, L., and Chen, D. (1996) Gene Exp. 4, 45–57
20. Higuchi, R. (1990) in PCR Protocols (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., eds) pp. 177–183, Academic Press Inc., San Diego
21. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1990) Current Protocols in Molecular Biology, units 12.1–12.3, Current Protocols, U. S. A.
22. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1989) Trends Biochem. Sci. 11, 20–25
23. Kingsley, C., and Winoto, A. (1996) Mol. Cell. Biol. 12, 4251–4261
24. Hagen, G., Muller, S., Beato, M., and Suske, G. (1992) Nucleic Acids Res. 20, 5519–5525
25. Halazonetis, T. D., and Kandil, A. N. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 6162–6166

[FIG. 13. Cis-elements and trans-factors regulating nonmuscle myosin II heavy chain-A gene. Nucleotide sequence of the N2d’ region (1–120) are shown. The sequence of the N2d region corresponds to the sequence from 11 to 110. Protein binding elements C, A, and F determined in this study are indicated. Transcriptional factors Sp1 and Sp3 bind to element C. The curved arrows represent a possible functional interaction.]
26. Gregor, P. D., Sawadogo, M., and Roeder, R. G. (1990) *Genes Dev.* 4, 1730–1740
27. Boyd, K. E., and Farnham, P. J. (1997) *Mol. Cell. Biol.* 17, 2529–2537
28. Blackwell, T. K., and Weintraub, H. (1990) *Science* 250, 1104–1110
29. Mitchell, P. J., and Tjian, R. (1988) *Science* 245, 371–378
30. Dennig, J., Beato, M., and Suske, G. (1996) *EMBO J.* 15, 5659–5667
31. Birnbaum, M. J., van Wijnen, A. J., Odgren, P. R., Last, T. J., Suske, G., Stein, G. S., and Stein, J. L. (1995) *Biochemistry* 34, 16503–16508
32. Liang, Y., Robinson, D. F., Dennig, J., Suske, G., and Fahl, W. E. (1996) *J. Biol. Chem.* 271, 11792–11797
33. Noti, J. D. (1997) *J. Biol. Chem.* 272, 24038–24045
34. Majello, B., De Luca, P., and Lania, L. (1997) *J. Biol. Chem.* 272, 4021–4026
35. Kennett, S. B., Udvardia, A. J., and Horowitz, J. M. (1997) *Nucleic Acids Res.* 25, 3110–3117
36. Seto, E., Lewis, B., and Shenk, T. (1993) *Nature* 365, 462–464
37. Gualberto, A., and Baldwin, A. S., Jr. (1995) *J. Biol. Chem.* 270, 19680–19683
38. Merika, M., and Orkin, S. H. (1995) *Mol. Cell. Biol.* 15, 2437–2447
39. Lin, S.-Y., Black, A. R., Kostic, D., Pajovic, S., Hoover, C. N., and Azizkhan, J. C. (1996) *Mol. Cell. Biol.* 16, 1668–1675
40. Lin, J.-X., and Leonard, W. J. (1997) *Mol. Cell. Biol.* 17, 3714–3722
41. Pascal, E., and Tjian, R. (1991) *Genes Dev.* 5, 1646–1656