Helix-Loop-Helix Proteins as Regulators of Muscle-specific Transcription

Diane G. Edmondson and Eric N. Olson*

From the Department of Biochemistry and Molecular Biology, University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Unraveling the mechanisms that control cell type-specific transcription represents a major challenge for molecular biologists. The genes expressed by a differentiated cell depend on the cell's unique developmental history and the integration of diverse extracellular signals that regulate intracellular pathways of signal transduction. Skeletal muscle provides an attractive system for analyzing the roles of cell identity and extracellular signals in the control of cell type-specific transcription. Myoblast differentiation is accompanied by the transcriptional activation of a large set of muscle-specific genes, many of which have been cloned and characterized. In addition, skeletal muscle cells can be maintained in tissue culture as established cell lines that recapitulate the events associated with myoblast differentiation in vivo. The program for muscle-specific transcription is also regulated by numerous growth factor and oncogenic signaling pathways, making this system amenable to investigations of the regulatory interactions between the cellular circuits that control cell proliferation and differentiation.

The recent cloning of the MyoD family of myogenic regulators, which can directly activate skeletal muscle-specific genes and convert a wide range of cell types into skeletal muscle, has provided insight into the mechanisms that regulate muscle-specific transcription and has revealed basic regulatory strategies that may control cell type-specific transcription in other specialized cell types (for reviews, see Refs. 1–3). While other cell type-specific transcription factors have been identified, members of the MyoD family are unique thus far in their abilities to orchestrate an entire program of cell differentiation and to activate and maintain their own transcription. This review will focus on the role of the MyoD family in the control of muscle-specific transcription, on the influence of extracellular signals and cell identity in influencing the actions of these factors, and on the relationship between this family of cell type-specific regulators and the factors that may control specific patterns of gene transcription in other cell types.

Regulation of Muscle-specific Transcription by the MyoD Family of Myogenic Regulators

Skeletal myoblasts arise from mesodermal stem cells that become restricted to a myogenic fate through mechanisms whose details remain obscure. Although undifferentiated and capable of continued proliferation, myoblasts are committed to the myogenic lineage and ultimately differentiate when they receive the appropriate environmental signals. Once myoblasts enter the differentiation pathway they withdraw from the cell cycle and fuse with neighboring myoblasts to form multinucleate myotubes. A battery of muscle-specific genes whose products are required for the unique contractile and metabolic properties of the muscle fiber are activated at the same time. Myoblast differentiation in tissue culture is inhibited by FGF, TGF-β, and other undefined serum components.

Studies throughout the early 1980s supported the notion that one or a few regulatory genes had the potential to initiate a cascade of events leading to activation of the muscle differentiation program. The existence of myogenic regulatory genes was verified first with the cloning of MyoD, which was identified by subtraction-hybridization of cDNAs representing transcripts expressed in undifferentiated myoblasts but not in 10T1/2 cells (5). Soon thereafter, the MyoD-related genes myogenin (6, 7), myf5 (8), and MRF4 (9), collectively referred to as the MyoD family, were isolated. Each of these factors is expressed exclusively in skeletal muscle; there is no detectable expression in cardiac or smooth muscle, despite the fact that all three muscle cell types express many of the same muscle genes. Forced expression of any one of these proteins in a variety of nonmuscle cell lines or primary cell types is sufficient to activate the complete program of myogenic differentiation when the cells are placed in a growth factor-deficient environment. In addition to activating the expression of muscle-specific genes associated with terminal differentiation, members of the MyoD family can activate one another's expression and positively autoregulate their own transcription in transfected cells (10, 11). These auto- and cross-regulatory interactions have been proposed to provide stability to the myogenic phenotype and to lead to amplification of the expression of these factors above a threshold necessary for commitment to terminal differentiation.

Different cell types vary in their susceptibility to myogenic conversion by members of the MyoD family. Aspects of myogenesis have been shown to be activated by these myogenic regulators in fibroblasts, neurons, adipocytes, chondrocytes, bone cells, smooth muscle cells, and melanocytes, indicating that muscle-specific genes are not irreversibly repressed in cells in which they would normally not be transcribed (reviewed in Refs. 1 and 12). In some of these cell types, members of the MyoD family activate the complete myogenic program, resulting in the formation of multinucleate myotubes with a fully assembled contractile apparatus, whereas in others, only subsets of muscle-specific genes may be transiently expressed. In some cell types, such as adipocytes, activation of the myogenic program by the MyoD family extinguishes the endogenous program of cell type-specific transcription, whereas in other cell types, such as melanoma and neuroblastoma, the myogenic program is co-expressed with the endogenous program (13). There are also several cell types that are refractory to myogenic conversion by members of the MyoD family. These include HeLa cells, liver cells, CV-1 kidney cells, and several lines established from the skeletal muscle tumor rhabdomyosarcoma (13–15). The failure of the myogenic regulators to activate myogenesis in these cell backgrounds may be due to the absence of cellular factors with which they normally collaborate to induce muscle transcription or the presence of inhibitory factors that interfere with their functions, or both.

The HLH Motif Is an Ancient Protein Motif Contained in Proteins That Regulate Cell Fate

Members of the MyoD family share about 80% homology within a segment of ~70 amino acids that encompasses a basic region followed by a predicted HLH motif, in which two amphipathic α-helices are separated by an intervening unstructured loop (reviewed in Ref. 16). Members of this regulatory gene family have been identified in a variety of vertebrates as well as in several invertebrates, including Drosophila, Caenorhabditis elegans, and sea urchin. Whereas the genomes of all vertebrates examined thus far encode multiple members of the MyoD family,
invertebrates appear to contain only a single member of the gene family. The myogenic factors from sea urchin and C. elegans can efficiently activate myogenesis in 10T1/2 cells (17, 18), indicating that the mechanism for muscle gene activation has been highly conserved.

The basic-HLH (bHLH) motif defines a superfamily of proteins that regulate cell type-specific transcription, as well as cell proliferation and transformation (16). Among these proteins are the products of the *Drosophila* *achute-scute* complex and their mammalian homologues, which regulate neurogenesis, twist, which regulates mesoderm formation, and daughterless, which participates in sex determination and formation of the peripheral nervous system in *Drosophila*. Members of the *myc* family of oncogene products and their dimerization partner Max also contain conserved bHLH motifs that are essential for cellular transformation. The bHLH motif is also found in a class of widely expressed factors known as E-proteins, which includes E12, E47, and HEB, as well as in USF and AP4, which are involved in general transcription (19).

Mutagenesis of HLH proteins has shown that the HLH motif serves as an interface for dimerization, which brings together the basic regions of HLH proteins to form a composite DNA-binding domain (20, 21). Recent studies suggest that HLH proteins dimerize to form a structure resembling a 4-helix bundle in which the helices are parallel (22) (Fig. 1). Members of the MyoD family do not efficiently homodimerize but readily form heterodimers with ubiquitous HLH proteins of the E-protein class (23). All bHLH proteins analyzed thus far recognize the nucleotide consensus sequence CANNTG, known as an E-box. This DNA sequence motif is found in the control regions of most skeletal muscle-specific genes, as well as several other cell type-specific genes (1–3).

Members of the MyoD family contain transcription activation domains in their amino and carboxyl termini that are important for efficient activation of muscle-specific transcription (Fig. 2). In contrast to the myogenic bHLH region, which functions in only certain cellular contexts (see below), the activation domains of the myogenic regulators, when fused to the DNA-binding domain of yeast GAL4, are active in cells that can and cannot be converted to muscle, indicating that they do not depend on cell-specific co-regulators for activity (2, 15, 24).

The basic regions of myogenic HLH proteins play a dual role in the control of muscle-specific transcription by mediating DNA binding and by conferring specificity to transcriptional activation. Mutagenesis of the basic region prevents DNA binding but not dimerization (21, 25). HLH proteins that lack functional basic regions have also been identified and shown to function as negative regulators of E-box-dependent transcription. One such protein is Id (inhibitor of differentiation), which dimerizes preferentially with E-proteins and thereby inactivates myogenic HLH proteins by sequestering their dimerization partners (26). The presence of Id at high levels in undifferentiated myoblasts and other cell types and its down-regulation during differentiation has led to the notion that it may play a role as a negative regulator of myogenesis and other programs for cell type-specific transcription.

The importance of the basic region for muscle-specific transcription was revealed by domain swap experiments in which the basic regions of MyoD and myogenin were replaced with those of other HLH proteins, such as E12 and the *achaute-scute* complex gene product (21, 25). The resulting chimeric proteins retain the ability to dimerize with E-proteins and bind DNA, but they are devoid of myogenic activity when assayed for their ability to activate endogenous or exogenous muscle-specific genes. The behavior of these basic domain mutants indicates that DNA binding, although necessary for activation of muscle transcription, is not sufficient and suggests that additional events mediated by the basic region are essential for activation of the myogenic program. Fine mapping of the basic regions of MyoD and myogenin has revealed that two adjacent amino acids, alanine-threonine, in the center of their DNA-binding domains confer muscle specificity to the basic region. These residues, which compose a so-called myogenic recognition motif, are conserved in the basic regions of all known myogenic HLH proteins and are absent from the basic regions of the more than 40 other HLH proteins described to date. The specificity and conservation of these residues suggest that they represent an essential component of an ancient mechanism for activation of the myogenic program.

The precise mechanism whereby the basic regions of myogenic HLH proteins direct muscle-specific transcription is unknown, but the behavior of basic domain mutants of these proteins is consistent with the possibility that this protein domain is required for interaction with a third protein, which has been referred to as a “co-regulator” or “recognition factor.” Since the alanine-threonine of the myogenic recognition motif is predicted to lie within the major groove of the DNA-binding site (27), these amino acid residues may not interact directly with such co-regulators, but they could conceivably alter the conformation of the proteins so that they could participate in such protein-protein interactions.

**Myogenic HLH Proteins Function within a Regulatory Circuit That Controls Muscle-specific Transcription**

E-boxes are present in the control regions of most skeletal muscle genes where they mediate muscle-specific transcription and trans-activation by myogenic HLH proteins. These E-boxes are generally surrounded by binding sites for other transcription factors that collaborate with the myogenic regulators to induce muscle transcription. The α-cardiac actin gene promoter, for example, is regulated by cooperative interactions between MyoD, serum response factor, and myoD1 (28), while the muscle creatine kinase gene is regulated by an enhancer that contains binding sites for myogenic HLH proteins, the muscle enhancer factor-2 (MEF-2), and the mesoderm-restricted homeodomain protein M-Hox (29).
There are also examples of muscle genes that are regulated by myogenic HLH proteins but which lack E-boxes in their control regions. Activation of these types of muscle-specific genes appears to be mediated by “intermediate” myogenic factors, which are themselves regulated by the MyoD family. Among these factors is MEF-2, a member of the MADS box family of transcription factors (30, 31), which is induced when myoblasts enter the differentiation pathway (32). MEF-2 binds an A+T-rich element found in the promoters and enhancers of numerous muscle-specific genes where it is important for transcriptional activity. MEF-2 activity is absent from nonmuscle cells but can be induced by myogenin or MyoD, suggesting that it functions “downstream” of these myogenic regulators in a regulatory pathway (33). Induction of MEF-2 activity by myogenin can occur in cell types that can and cannot be converted to muscle, suggesting that MEF-2 is regulated independently of the genes involved in terminal differentiation.

Paradoxically, MEF-2 binds the myogenin promoter, and its binding site is required for full activity of that promoter (11). The ability of myogenin to induce MEF-2, which in turn regulates myogenin transcription, provides a mechanism whereby these factors can amplify and maintain one another’s expression when myoblasts are triggered to differentiate. MEF-2 is also expressed in cardiac and smooth muscle, which do not express known members of the MyoD family. The factors responsible for MEF-2 induction in these muscle cell types remain to be identified.

Members of the MyoD Family Exhibit Distinct Patterns of Expression in Vivo and in Vitro

Skeletal muscle in vertebrates arises in the somites, which form by segmentation of the paraxial mesoderm along the neural tube. The four members of the MyoD family are first expressed in the somites when myogenic precursors appear and subsequently are present in the limb bud during muscle fiber formation (34). Although members of the MyoD family can positively autoregulate one another’s expression in transfected cells and all share the property of being expressed specifically in skeletal muscle cells, each gene is activated at a slightly different time during embryogenesis, suggesting that they respond to distinct regulatory inputs.

Members of the MyoD family also show unique patterns of expression during myogenesis in tissue culture. Either MyoD or myf5 is generally expressed in proliferating myoblasts prior to initiation of differentiation, whereas myogenin does not become expressed at high levels until differentiation has been triggered by withdrawal of growth factors. MRF4 is rarely expressed in muscle fibers in tissue culture. In contrast to MyoD and myf5, which are expressed only in subsets of skeletal muscle cells, myogenin is expressed during differentiation of all skeletal muscle cells examined thus far, suggesting that the myogenin gene responds to a common myogenic regulatory pathway or that myogenin is essential for skeletal muscle differentiation.

Why Are There Multiple Myogenic HLH Proteins?

If each member of the MyoD family can activate the complete myogenic program, why are there four myogenic regulatory genes? It has been difficult to determine whether members of the MyoD family function in the same or different developmental pathways. However, studies have shown that members of the MyoD family can activate one another’s expression and because most assays for their functions involve overexpression in transfected cells, which may mask subtle functional differences. There is evidence, however, that MyoD is important for inducing myoblast fusion. The BC3H1 muscle cell line expresses myogenin, but does not express MyoD, and cannot form myotubes. Forced expression of MyoD in BC3H1 cells rescues the ability of the other’s expression when activation experiments also indicate that MRF4 differs from myogenin and MyoD in its ability to activate certain muscle-specific genes (36). Whereas these three myogenic regulators bind with equivalent affinities to the E-boxes in the muscle creatine kinase, troponin-I, and myosin light chain-1/3 enhancers, MRF4 fails to activate transcription of these genes. Swapping of the amino and carboxyl termini of myogenin and MRF4 exchanges their specificities for muscle creatine kinase enhancer activation (37). These results suggest that myogenin and MRF4 can discriminate between muscle-specific enhancers and that target gene specificity of trans-activation is determined by nonconserved protein domains surrounding the bHLH region. This type of transcriptional specificity could provide the basis for the unique patterns of muscle-specific gene expression in different muscle fiber types, depending on the combinations of myogenic regulators that are expressed.

Gene Targeting Experiments Yield Unanticipated Results

A direct means of determining the functions of myogenic HLH proteins during embryogenesis and of assessing the extent of their redundancy is to inactivate these genes within the whole organism. Such studies have recently been undertaken and have yielded unanticipated phenotypes. Deletion of the MyoD gene from C. elegans (18) leads to lethality and defects in muscle organization but does not prevent expression of muscle-specific genes. The ability of muscle genes to be expressed in the absence of MyoD raises the possibility that experiments performed in cell lines such as 10T1/2 may have led us to exaggerate the roles of myogenic HLH proteins and overlook other pathways for activation of muscle-specific transcription. Activation of myogenesis in the absence of MyoD is most easily explained by redundant myogenic regulators for activation of the myogenic program or by additional members of the MyoD family that have diverged substantially and have yet to be detected.

The MyoD and myf5 genes have also been inactivated in mice by homologous recombination, with surprising results. Mice lacking MyoD have normal skeletal muscle, are fully viable, and do not exhibit obvious defects (38). The only molecular difference noted thus far is that the level of myf5 is slightly elevated relative to wild-type mice. Mice lacking myf5 show an unanticipated phenotype, in which their ribs are either totally lacking or are severely truncated (39). These animals die at birth because they cannot breathe; however, the muscle of myf5-null mice is normal. The presence of muscle in these animals suggests that there is redundancy of functions of myogenic HLH proteins in vertebrates or that these proteins are nonessential for muscle formation in vivo. Analysis of myogenesis knockouts is under way and has revealed that mice lacking myogenin have grossly abnormal muscle, in which few myofibers form, and expression of many muscle-specific genes is severely reduced.2 These animals are alive at birth but never move or breathe and rapidly die. These results suggest that myogenin is essential for complete formation of functional muscle in vivo.

Mutual Antagonism between the Cellular Circuits That Control Myoblast Proliferation and Differentiation

Whether a myoblast differentiates or divides is dictated by a balance of opposing cellular signals controlled by myogenic HLH proteins and peptide growth factors. Thus, when myoblasts are exposed to high concentrations of growth factors, myogenic HLH proteins are inactivated such that they cannot activate muscle-specific genes. Conversely, overexpression of the myogenic factors can induce cell cycle withdrawal and initiate myogenesis even when cells are faced with high concentrations of mitogens or activated oncogenes (40–42). The ability of growth factors, acting at the cell surface, to regulate the activity of myogenic HLH proteins in the nucleus suggests that these proteins may serve as end points for intracellular growth factor signal transduction cascades. However, the signal transduction pathways through which growth factors inhibit muscle-specific gene activation are only beginning to be understood.

Recently, it was shown that activated protein kinase C, which transduces mitogenic signals, can substitute for exogenous peptide growth factors and can silence the transcriptional activities of myogenic HLH proteins (43). Phosphopeptide mapping of myogenin in transfected cells revealed that the threonine residue 60 in the myogenic recognition motif of the basic region (Fig. 2) is a protein kinase C phosphorylation site and that when phosphorylated by protein kinase C, this site mediates repression of myogenin’s activity through a loss in DNA binding activity. FGF

---

1. Hasty, J. H. Morris, D. Edmondson, J. Venuti, A. Bradley, E. Olson, and W. Klein, manuscript in preparation.
induces phosphorylation of the same site and represses myogenin's transcription activating potential, whereas a myogenin mutant lacking the protein kinase C phosphorylation site fails to be repressed by FGF. The presence of this phosphorylation site in the DNA-binding domains of all known myogenic HLH proteins suggests that it represents a conserved target for negative control of these proteins' functions via regulated phosphorylation.

The cAMP signal transduction pathway, mediated by the cAMP-dependent protein kinase catalytic (C)-subunit, also can silence the activity of myogenic HLH proteins (44). Myogenin has been shown to be a direct substrate for the C-subunit of protein kinase A. However, mutations within the protein kinase A sites do not affect myogenin's activity, suggesting that inhibition by the C-subunit involves an indirect mechanism with one or more intermediate steps.

TGF-β inhibits the activity of myogenic HLH proteins through a mechanism independent of DNA binding and mediated by the bHLH region (45, 46). The behavior of myogenic HLH proteins in cells exposed to TGF-β is similar to that of basic domain mutants that bind DNA but do not activate transcription. Repression by TGF-β is consistent with a model in which a TGF-β-dependent signaling pathway interferes with the expression or activity of a co-regulator that recognizes the myogenic basic region.

Several indirect mechanisms for growth factor-dependent inactivation of myogenic HLH proteins have been described. In addition to the induction of IκB, which forms inactive heterodimers with E-proteins, c-Jun and c-Fos can block muscle gene expression in response to mitogenic signals. Transcriptional repression by c-Jun can be mediated by direct interaction between the leucine zipper of c-Jun and the HLH region of MyoD (47). The amino-terminal activation domain of c-Jun also can inactivate myogenic HLH proteins through a mechanism that appears to involve competition for the putative co-regulator that recognizes the myogenic basic region (48). The mechanism through which Fos blocks the activities of myogenic HLH proteins is unknown.

**Perspective**

Where do members of the MyoD family function within the regulatory cascade leading to myogenesis? The ability of these proteins to activate the morphological and molecular events of myogenesis in nonmuscle cell types illustrates that these gene products can alter cell fate and orchestrate an entire program of cell type-specific transcription, and it is consistent with the notion that these genes function as determination genes that mediate commitment to the myogenic lineage. However, as their properties have been analyzed in greater detail, certain aspects of this function have become more problematic in defining the exact role of these proteins in establishment of the myogenic phenotype. In particular, their ability to inhibit cell proliferation provides a property that might be expected of a transcription factor involved in terminal differentiation rather than cell determination. Moreover, the expression of myogenic HLH proteins can be extinguished in response to various extracellular and intracellular signals without altering the position of myoblasts in the lineage, suggesting that other factors may mediate lineage commitment. Finally, the observation that MyoD can be deleted from the genome of at least certain species without a loss in muscle-specific transcription suggests the existence of additional regulatory proteins that act independently of the MyoD family to induce muscle transcription. We therefore view these proteins as muscle-specific transcription factors that participate in terminal differentiation. According to this view, these proteins would function within a regulatory cascade that can be entered at multiple levels. Thus, forced expression of these proteins in transfected cells can activate the “downstream” events of myogenesis and leads to the same ultimate phenotype as would be achieved by activating more “upstream” events associated with lineage determination. The apparent redundancy in the system may be mediated by other intermediate regulatory proteins which are known to activate muscle-specific transcription and may themselves be induced by regulatory pathways in addition to those under control of the MyoD family.

In the future, it will be especially interesting to further define the potential mechanisms for activation of muscle-specific transcription in the absence of the MyoD family, to unravel the mechanisms that regulate the expression of myogenic HLH proteins themselves, and to further elucidate the roles of cell identity and signal transduction in governing the actions of these factors.

**Acknowledgment**—We appreciate the secretarial assistance of K. Tucker.