18. D. Johnson, *Nature* **335**, 155 (1987).
19. M. Padberg and G. Rinaldi, *Oper. Res. Lett.* **6**, 1 (1987).
20. _____, Report R. 247 (Instituto di Analisi Dei Sistimi ed Informatica del CNR, Rome, 1988).
21. M. Grötschel and O. Holland, Report No. 73 (Institut für Mathematik, Universität Augsburg, Germany, 1988).
22. P. C. Kanellakis and C. H. Papadimitriou, *Oper. Res.* **28**, 1086 (1980).
23. O. B. G. Madsen, *J. Oper. Res. Soc.* **39**, 249 (1988).
24. R. D. Plante, T. J. Lowe, R. Chandrasekaran, *Oper. Res.* **35**, 772 (1987).
25. M. Fischetti and P. Toth, "An additive bounding procedure for the asymmetric traveling salesman problem" (DEIS, University of Bologna, Italy, 1989).
26. J. F. Pekny and D. L. Miller, *Math. Program.*, in press.
27. _____, D. Stodolsky, *Oper. Res. Lett.*, in press.
28. D. Stodolsky, J. F. Pekny, D. L. Miller, *Eng. Des. Res. Cent. Rep. 05-25-88* (Carnegie Mellon Univ., Pittsburgh, PA 1988).
29. D. L. Miller and J. F. Pekny, *Oper. Res. Lett.* **8**, 129 (1989).
30. E. Balas, D. L. Miller, J. F. Pekny, P. Toth, *J. Assoc. Comput. Mach.*, in press.
31. R. E. Tarjan, *Data Structures and Network Algorithms* (Society for Industrial and Applied Mathematics, Philadelphia, 1983).
32. R. M. Karp, *SIAM J. Comput.* **8**, 561 (1979).
33. Solution of the root vertex AP followed by patching with $\lambda = \infty$.
34. M. Bellmore and J. C. Malone, *Oper. Res.* **19**, 278 (1971).
35. C. H. Papadimitriou and K. Steiglitz, *ibid.* **26**, 434 (1978).
36. $c_{ij} \leq c_{ik} + c_{kj}$ for all $i, j, k \in V$.
37. We thank Cray Research Inc. for computer access and technical support.

The *myoD* Gene Family: Nodal Point During Specification of the Muscle Cell Lineage

HAROLD WEINTRAUB, ROBERT DAVIS, STEPHEN TAPSCOTT, MATTHEW THAYER, MICHAEL KRAUSE, ROBERT BENEZRA, T. KEITH BLACKWELL, DAVID TURNER, RALPH RUPP, STANLEY HOLLENBERG, YUAN ZHUANG, ANDREW LASSAR

The *myoD* gene converts many differentiated cell types into muscle. MyoD is a member of the basic-helix-loop-helix family of proteins; this 68–amino acid domain in MyoD is necessary and sufficient for myogenesis. MyoD binds cooperatively to muscle-specific enhancers and activates transcription. The helix-loop-helix motif is responsible for dimerization, and, depending on its dimerization partner, MyoD activity can be controlled. MyoD senses and integrates many facets of cell state. MyoD is expressed only in skeletal muscle and its precursors; in nonmuscle cells *myoD* is repressed by specific genes. MyoD activates its own transcription; this may stabilize commitment to myogenesis.

THE *myoD* GENE IS CAPABLE OF ACTIVATING PREVIOUSLY silent muscle-specific genes when introduced into a large variety of differentiated cell types with a viral long terminal repeat (LTR) used to promote constitutive transcription (*1–4*). In certain cell types, the entire program for muscle differentiation seems to be activated (*3*). The range of cell types converted to muscle by *myoD* includes a number of fibroblast cell lines, adipocytes, melanoma cells, a hepatoma cell line, neuroblastoma cells, osteosarcoma cells and P19 teratocarcinoma cells, as well as primary cultures of chondrocytes, smooth muscle, retinal pigment, fibroblasts, and brain cells. MyoD is expressed only in skeletal muscle. Cardiac and smooth muscle, which express many of the same muscle-specific structural genes as skeletal muscle, do not express MyoD (*1, 5*). The MyoD protein seems to activate myogenesis by directly binding to the control regions of muscle-specific genes (*6*). On the basis of these properties, we refer to *myoD* as a "master regulatory gene." Implicit in this shorthand is the fact that other factors must be responsible for the initial activation of *myoD*, and that the activity of the MyoD protein, itself, could be and is regulated. We view *myoD* as a "nodal point" (*2*) in the flow of myogenic information from the early embryo to the mature myofiber and, as discussed below, consider all members of the *myoD* family (*myogenin*, *myf-5*, *mrf-4-herculin*) to perform more or less the same "function," because assays to date have not dramatically distinguished one from another. In contrast to segmentation genes, homeotic genes, lineage genes, and the like, studied in *Drosophila* or *Caenorabditis elegans*, *myoD* seems to affect the identity of a single cell type, not constellations of many types of cells.

We describe here the structure of MyoD; how it activates the myogenic program; and how *myoD*, itself, is transcriptionally and posttranscriptionally regulated during development. The story can appear extremely simple; however, not unexpectedly, new findings bring new paradoxes and complexities, which we will also explore.

The authors are at the Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104.

A Single Genetic Function Can Activate Myogenesis

The notion that myogenesis can be controlled by a single master regulatory gene originates in a series of experiments from Holtzer's lab (7). When the thymidine analog bromodeoxyuridine (BrdU) was incorporated into DNA, myogenesis in culture was inhibited, but, with continued growth in the absence of BrdU, the reappearance of muscle was very rapid. This suggested that incorporation of BrdU into DNA inhibited one or a few targets on one or a few chromosomes and, in turn, these targets were capable of reactivating the myogenic program when subsequently segregated to daughter cells after growth and division in the absence of BrdU (*8*). We now know that BrdU turns off transcription of *myoD*, although we do not know how (*9*).

The idea of a pivotal control gene gained momentum from the work of Taylor and Jones (*10*), who showed that brief treatment of C3H $10T\frac{1}{2}$ fibroblastic cells with 5-azacytidine induced the formation of large numbers (25 to 50%) of myogenic colonies, as well as fewer numbers of chondrogenic and adipogenic colonies. The high frequency of myogenic colonies suggested that 5-azacytidine, which

is incorporated into DNA and inhibits subsequent methylation of cytosine, resulted in the activation of only one or a few target genes and that this activation was heritable (*10, 11*). The *myoD* gene is highly methylated in C3H 10T½ cells and becomes demethylated in 5-azacytidine–derived muscle lines (*12*). Genomic DNA transfection experiments have shown that the muscle phenotype can be conferred to recipient C3H 10T½ cells by transfection with DNA from 5-azacytidine–induced myoblasts, but not with DNA from parental C3H 10T½ cells (*13*). Because the probability of transferring two loci in this type of experiment is so low, these results indicate that a single genetic locus can activate the myogenic program in C3H 10T½ cells and that the DNA at this locus is modified, presumably by methylation, in C3H 10T½ cells, but not modified in myoblasts. Whereas methylation of *myoD* seems to be important in silencing the *myoD* gene in nonmuscle tissue culture cell lines (*12*), in primary nonmuscle cells the gene is not methylated and is probably repressed by trans-acting factors (see below). The *myd* gene may be responsible for the myogenic phenotype observed after genomic transfections and may activate *myoD* (*13*). However, the relation between *myd* and *myoD* is unclear because *myd* has not yet been cloned and characterized.

MyoD is a transcription factor that activates muscle-specific genes. The *myoD* gene was initially cloned by subtractive hybridization in an attempt to identify messenger RNAs (mRNA) that were expressed in proliferating myoblast cell lines, including 5-azacytidine–induced muscle lines, but not expressed in parental C3H 10T½ cells or in mutant muscle lines that had lost the ability to differentiate (*1*). MyoD is a nuclear protein, 318 amino acids in length, that binds to many muscle-specific enhancers. Although it is phosphorylated, the function of the phosphorylation has not been defined. The protein and RNA turn over rapidly. MyoD binds to a consensus DNA binding sequence that includes a CA– –TG sequence present in most muscle-specific enhancers (*6*). The apparent simplicity of this sequence belies a rather sophisticated interplay between protein, DNA, and subsequent cell type–specific transcriptional activation as described below. Many, if not most, muscle enhancers contain multiple MyoD binding sites (*14, 15*).

The *myoD* gene is a member of a multigene family of myogenic determination genes [*myogenin* (*16*), *myf-5* (*17*), *mrf4-herculin-myf-6* (*18*)], all of which have largely been defined by their ability to convert C3H 10T½ fibroblasts to muscle. At present, it is not clear whether these genes have biologically distinct functions, although quantitative differences in activity and time of embryonic expression are detectable. Moreover, although there are exceptions, each of these genes can activate the others (and the corresponding endogenous copy of itself) after it is transfected into a recipient cell (*19, 20*). This auto- and cross-activation network leads to the production of large amounts of active, myogenic regulatory proteins, once any one of these genes is activated. Thus, in principle, multiple cross-activating myogenic genes can be used to reinforce or commit myoblasts to myotube production, so that once the initial switch or decision to become muscle is made, for example, by activating any one of these genes, it is usually irreversible.

Only 68 amino acid residues of MyoD are required for stable myogenic conversion of C3H 10T½ cells (*21*). This region contains the DNA binding domain, which is a putative basic helix-loop-helix structure (bHLH) (*22, 23*). The structure of the bHLH motif is not known but only inferred from sequence data and genetic mutation experiments. There are now well over 30 known bHLH proteins, sometimes referred to as the *myc* homology family (*24*). Extensive mutagenesis (*23, 25*) has revealed that the HLH region is required for dimerization and that the adjacent basic region is required for DNA binding. Because very minor alterations at the junction between the basic and HLH region inhibit DNA binding but not dimerization, it is thought that the HLH domains (*23*) orient the two basic regions contributed by the individual monomers. Whether the basic region itself takes on a specific structure when bound to DNA is not known, although for MyoD inspection suggests an α helix is permissible. Methylation interference indicates that the basic region interacts with the major groove (*6*). Whereas the loop is highly conserved in size and sequence in MyoD molecules from humans to nematodes, a large number of mutations introduced into the loop have, as yet, failed to indicate a specific function for the loop. It may serve as a spacer between the two putative helices of the HLH domain so that helix 1 can fold back and interact with helix 2 (*23*). Recently, the NH_2-terminal 50–amino acid residues of MyoD have been shown to have an activating domain (*26*) and also to mediate cooperative (*27–29*) binding between adjacent MyoD binding sites. No function has yet been attributed to the remainder of the molecule, including the COOH-terminal 160–amino acid residues.

Whereas MyoD can bind DNA as a homodimer, it binds to its target sequence more than ten times better as a heterodimer with one of the differentially spliced bHLH proteins encoded by the *E2A* gene, first identified by Murre *et al.* (*28*). Sun and Baltimore (*29*) have shown that the increase in overall binding affinity of the heterodimer is attributable to an increase in dimerization, indicating that under certain conditions dimerization can be limiting. The *E2A* gene is expressed in many different cell types, including muscle. Like MyoD and Myf-5, it contains an activation domain as demonstrated by standard fusions to a Ga14 DNA-binding protein fragment that has no activation activity by itself (*30*). There are three indications that *E2A* may be required for MyoD activity in muscle: (i) Factors extracted from muscle cells that can bind to the enhancer of the muscle-specific creatine phosphokinase (MCK) gene include a MyoD-E2A (or Myogenin-E2A) hetero-oligomeric species. (ii) C3H 10T½ cell lines expressing antisense E2A RNA fail to express E2A protein and fail to convert to muscle after either 5-azacytidine treatment or infection with a *myoD* retrovirus. (iii) Overexpression of MyoD from a replicating plasmid in Cos cells fails to efficiently activate muscle-specific genes in an expression vector; however, when excess E2A is also provided, maximal activation is observed (*31*).

In order to activate muscle-specific genes, MyoD complexes must bind to two or more upstream sites (*27–29*). As assayed by dissociation rate, MyoD binds cooperatively, mediated by its NH_2-terminal acidic domain, which also harbors the activation function. Cotransfection of a MyoD expression vector with a muscle-specific expression vector carrying a single MyoD binding site results in no activity. Cotransfection with a vector carrying two sites leads to high levels of activity; however, if a third plasmid carrying a single site is also included and is in excess, inhibition occurs. This suggests that plasmids carrying a single site will compete specifically for the activation of a vector with two sites. This shows that binding to a single site is not sufficient for transcriptional activation and that, like other activator proteins (*32*), MyoD must occupy two or more sites before transcription is initiated. Under certain circumstances, a single MyoD binding site can activate transcription if binding sites for other constitutively expressed transcription factors are also present and occupied. Thus, a single MyoD binding site will activate a human cardiac actin gene if adjacent Sp1 transcription factor and SRF (serum response factor) sites are also provided (*33*). A requirement for filling two or more sites might help to prevent incorrect activation when a single MyoD complex binds to nonspecific sites. In addition, because MyoD can bind cooperatively to DNA containing two sites, cooperativity may allow an all-or-none response to occur when amounts of active MyoD are beginning to accumulate after induction of myogenesis.

Why is myogenic activation specific? The CA– –TG motif present in

most muscle-specific enhancers is also present in tissue-specific enhancers for a variety of other cell types and in several constitutive enhancers. Moreover, many bHLH proteins, either expressed in all tissues (such as E2A) or expressed in specific tissues, often can bind to the same consensus sequences. A major paradox therefore is why MyoD activates only muscle-specific genes and why muscle-specific genes are only activated by MyoD. For example, one particular set of bHLH proteins encoded by the *achaete-scute* gene complex of *Drosophila* can also bind to the MyoD binding site (*22*) (as can homodimers of the protein products of the *E2A* gene). Moreover, both MyoD and the *achaete-scute* protein products can bind to consensus sites within the immunoglobulin (Ig) enhancer. The *achaete-scute* gene, like *myoD*, also seems to be a master regulator but for neurogenesis (*34*). Genetic evidence suggests that *achaete-scute* interacts with a second gene, *daughterless* (*35*), which is homologous to *E2A* and which is also expressed in many cell types.

If the *achaete-scute* protein products can bind to MyoD binding sites, why do they not activate muscle genes in nerve and, by analogy, why does MyoD not activate nerve-specific genes in muscle? It could be that additional nerve-specific or muscle-specific factors are required or alternatively that negative factors might prevent activation in the inappropriate cell. These proposals are restricted by the observations that forced expression of MyoD from an LTR can lead to the activation of endogenous muscle genes in many cell types, including neuroblastoma cells and primary nerve cells, and that cotransfected LTR-MyoD can activate a reporter vector containing a simple, duplicated MyoD binding site in a variety of cell types. These results suggest that myogenic activation does not usually require other tissue-restricted factors or factors not able to be activated by MyoD. That negative factors prevent activation of myogenic genes by bHLH proteins other than those represented by the MyoD family is appealing but to date is unsupported. Such negative factors would need to be recessive to MyoD to explain the ability of transfected MyoD to activate myogenesis in so many cell types.

A third possibility (which does not exclude the others) is suggested by a variety of MyoD mutations (*23*). Domain-swapping experiments with *achaete-scute* showed that the helix 1, loop, or helix 2 regions could be exchanged with MyoD without loss in the ability of the chimeric MyoD to form dimers with E2A, to bind to muscle-specific enhancers, or to activate myogenesis in C3H $10T\frac{1}{2}$ cells. However, when the basic region of the *achaete-scute* protein was transferred into MyoD, the chimeric protein could bind specifically to muscle-specific enhancer DNA and dimerize with E2A, but it failed to induce muscle-specific differentiation or to trans-activate a muscle-specific reporter gene. The same result occurs when the E2A basic region is introduced into MyoD. The E2A basic region differs from MyoD in only four of ten positions that are conserved among the MyoD family of myogenic regulatory proteins. Changing just one of these residues back to the corresponding MyoD amino acid, an alanine, restores DNA binding of the hybrid so that it is indistinguishable from wild-type MyoD (the original domain swap has about half the binding affinity), but myogenic activation still fails to occur. Changing a second amino acid, back to a threonine, restores myogenic activation (*36*). These results show that the specificity for myogenic activation resides in the basic region of MyoD and have led to the suggestion that activation requires properties of both the DNA sequence and the basic region of the protein. This idea suggests that if MyoD were to bind to an *achaete-scute* DNA target, transcriptional activation would not occur because it would be the wrong DNA sequence. With activity restricted to only the site having the proper DNA sequence, nonspecific effects are reduced; thus, specificity is not restricted only to DNA binding (which is limited by the extent to which the interacting protein and DNA surfaces can be in contact) but also to subsequent activation. Perhaps binding to the correct DNA sequence, but not the incorrect DNA sequence, releases what seems to be a "cryptic" activation domain at the NH_2-terminus of MyoD (*26*). Such a mechanism would limit competition among soluble transcriptional activating proteins.

The theme that activation of transcription requires a specific interplay between closely juxtaposed DNA and protein residues has also emerged with several other types of DNA binding proteins. For example, mutations in the DNA binding domain of steroid receptors can result in a protein that binds specifically to DNA but fails to activate transcription (*37*). A related result is that both Oct-1, which is ubiquitously expressed, and Oct-2, which is B cell–specific, can bind to the Ig enhancer, but only Oct-2 is capable of activation of transcription (*38*). On the other hand, there are also examples of variant DNA binding sites that bind transcription factors equivalently but fail to support activation (*39*). Although far from proven, there is increasing evidence that, as a result of a specific interaction between protein and DNA, a conformational change occurs that leads to a functionally active protein. If protease sensitivity is used as an assay, different conformations of the yeast *mcm-1* gene product are detected when it binds to an α cell–specific enhancer versus an *a* cell–specific enhancer (*40*).

The activity of the MyoD protein is regulated. Proliferating myoblasts in tissue culture express MyoD RNA and protein, yet do not differentiate (*1*). As assayed by in vivo footprinting, the MyoD binding site in the enhancer of the MCK gene is occupied in differentiated myotubes but not in myoblasts (*41*). Thus, there are physiological controls that regulate MyoD activity. The decision to either proliferate as myoblasts or differentiate into myotubes is governed by the presence or absence of serum. How MyoD activity is modulated by serum is not known. One attractive model is that a second type of HLH protein controls MyoD. By screening a complementary DNA library for genes whose protein products were homologous to helix 2, the *Id* gene (inhibitor of differentiation) was cloned and then characterized (*24*). It contains an HLH motif, but no basic domain, and is expressed in most cell types. The Id protein can form heterodimers with MyoD or E2A proteins, but these dimers fail to bind DNA; the behavior of Id is similar to that of a specific class of dominant negative mutations of MyoD that are mutated in the basic region or are missing this region entirely. These mutations inhibit the activity of wild-type MyoD, presumably by binding E2A but failing to allow DNA binding. Id inhibits the ability of MyoD to trans-activate a muscle-specific reporter gene, and, in stable muscle lines, forced expression of Id from a viral LTR retards myogenic differentiation (*42*).

Id binds about fivefold better to E2A than to MyoD. When added to MyoD-E2A–DNA complexes at 37°C, Id rapidly causes dissociation, suggesting that the equilibrium between these complexes can be reached very rapidly, within minutes. Expression of Id mRNA in myoblasts is decreased to a tenth or a twentieth of its original value by removal of serum, suggesting that Id transcription responds, directly or indirectly, to SRF and that *Id* may be controlled in much the same way as *fos* and *jun* or be controlled by them. One model proposes that, when *Id* transcription ceases, bound E2A is released and becomes free to form a complex with MyoD. This MyoD-E2A complex would be the active species in turning on transcription of muscle genes. Whether there is a separate role for putative MyoD homodimers is not known; nor can it be excluded that other HLH proteins not yet identified also interact along with MyoD or E2A to control additional cellular activities.

Id expression is also down-regulated when a number of other cell types are induced to differentiate, suggesting that they too are induced to differentiate by bHLH proteins. It is likely that *Id* expression is influenced by many factors and therefore monitors

whether conditions are appropriate for irreversible commitment to an overtly differentiated phenotype. Because some agents (such as transforming growth factor β1) that inhibit myogenesis do not increase Id RNA levels (*42, 43*) it seems likely that, in addition to control of HLH dimerization of MyoD being mediated by proteins such as Id, the activity of complexes containing the various members of the MyoD family can also be regulated by posttranslational protein modification or by association with other regulatory elements. These aspects of MyoD activity have not been fully explored.

Information Processing by HLH Proteins

An analog of *Id* has been known for many years in *Drosophila*. This gene, *emc* (extra macrochaetae), has recently been cloned (*44*) and is similar to *Id* in having a related HLH domain without a basic region. Loss-of-function mutations in *emc* result in extra bristles on the surface of the fly, which reflects the underlying differentiation of sibling cells to sensory neurons (*45*). Like *Id*, *emc* is expressed in most cells, although the obvious effects of available mutations occur in the nervous system. The *achaete-scute* gene is required for mutated *emc* to produce extra bristles. Extra bristles are also produced by overexpression of *achaete-scute*, but not when *emc* is also proportionately overexpressed. The two genes behave as if the EMC protein is a negative regulator of *achaete-scute* that titrates either the *achaete-scute* protein product or a dimerization partner of the *achaete-scute* product, such as Da (*34*).

The ability of HLH proteins to regulate expression by a process of protein:protein titration seems to have been adopted to count the number of X chromosomes for sex determination and dosage compensation in *Drosophila* (*46*). One of the transcripts of the *achaete-scute* gene complex, which is X-linked, seems to be the same gene as *sisterless b*, a locus that, when mutated, results in a failure to recognize the presence of the corresponding X chromosome and hence a failure to activate female-specific genes. A related phenotype is also observed in *da* mutants. Da is a constitutively expressed bHLH protein similar to E2A, and recently it has been shown to also be required for neurogenesis and myogenesis. In the central nervous system, *achaete-scute* mutations prevent the initial differentiation of peripheral neurons, and *da* seems to function at a slightly later stage, after the neurons are formed (*47*). The *achaete-scute* protein product homodimers may serve to initiate activity, whereas heterodimers formed from the products of the *achaete-scute–da* genes may maintain differentiation of neurons in the central nervous system. Although MyoD homodimers, E2A homodimers, and MyoD-E2A heterodimers all require the consensus CA– –TG site for binding, each has different specificities when the two internal base pairs and the flanking sequences are varied (*48*). How these different sequences affect enhancer recognition and subsequent transcriptional activation is not known. Similarly, specific instances where altering pairing partners alters half-site recognition in vivo have yet to be described.

Several types of biologically important decisions use information provided by HLH proteins: MyoD and the *achaete-scute* product seem to provide information about cell type; Id and EMC, a pathway for negative control; and E2A and Da, a presumed effector or integration function yet to be fully defined. These types of inputs then seem to be integrated by HLH interactions between pairs of individual monomers, and the result seems to be a specific binding capacity. For protein complexes containing Id and EMC, the choice is probably on-or-off; for complexes of E2A or MyoD homodimers or heterodimers, the result can be altered DNA binding specificity. Although it is clear that interactions between HLH proteins are very precise and specific, the rules are not yet known. If heterodimer formation with E2A is the limiting event, then it is possible that the process can become mutually exclusive; that is, if MyoD is complexed with E2A, then, for example, a small amount of aberrantly produced *achaete-scute* product will not be able to heterodimerize with E2A, and there is less danger of a muscle cell suddenly becoming a nerve cell. This type of competition may explain why, when overexpression of MyoD is forced in a nonmuscle cell, the endogenous program is often turned off (*2, 3*).

Once a decision to activate MyoD expression and function is made, it usually is irreversible because active MyoD turns on not only its own synthesis but also that of the other myogenic regulatory genes. Other factors not well understood also seem to contribute to the commitment to the myogenic state (*21*). A positive feedback loop can also be susceptible to interruption; thus, introduction of an overexpressing Id vector could inactivate MyoD, dampen positive feedback, and cause a cell to revert to a nonmyogenic phenotype. Similarly, other HLH proteins or factors that inactivate or compete with the dimerization or activation domain of MyoD might also do the same. For example, when MyoD-encoding mRNA is injected into two–cell stage frog embryos (*49*), muscle-specific actin is transiently turned on inappropriately. Nevertheless, the embryo develops normally. These results demonstrate that the MyoD mRNA product is not sufficient to irreversibly commit a cell to myogenesis. In contrast, the continued synthesis of MyoD mRNA from a DNA vector transfected into many tissue culture cells seems to be adequate to cause the cell to commit to myogenesis. The frog embryo experiments suggest that the presence of the MyoD mRNA in a developing embryo is not sufficient to activate a stable positive feedback loop. It is also possible that the presence of other HLH proteins activated by the normal differentiation process may limit the ability of MyoD protein to function fully, perhaps by competition for E2A proteins.

MyoD is controlled by and can control the state of cell growth. Holtzer and colleagues (*50*) first noted that a variety of transforming agents, in particular, the transforming gene *src*, would specifically inhibit myogenic differentiation. Subsequently, the list has been greatly extended to include genes such as *ras*, *fos*, *jun*, *fps*, *erbA* (*51*, and references therein), *myc* (*52*), and *E1A* (*53*), as well as chemical agents such as butyrate (*54*) and phorbol esters (*55*). Most of these reagents can inactivate the expressed MyoD protein; in addition, several (such as *ras* and *fos*) also inhibit *myoD* transcription. Whether this is a secondary effect due to an inhibition of the autoactivation function of MyoD protein or a more direct inhibition of *myoD* transcription remains to be determined. Rhabdomyosarcoma cells (derived from tumors of patients who harbor a genetic predisposition to myogenic tumors that maps near *myoD* on chromosome 11) differentiate poorly but express MyoD, suggesting that loss of antioncogene activity at the rhabdomyosarcoma locus can also impinge on MyoD function (*56*). The specific pathway by which each of these oncogenes, antioncogenes, and growth factors inhibits myogenesis provides a potential clue as to how MyoD might integrate information coming from many aspects of cellular function. Although we study the effects of these oncogenes and growth factors in tissue culture cells, this may only be a reflection of their true role during development.

The varied and complex effects of oncogenes on MyoD expression and activity may contribute to the failure of some cells to activate myogenic genes after introduction of MyoD (*2*). CV1 cells, HepG2 cells, and HeLa cells can express MyoD from an LTR-driven vector, but they fail to activate muscle-specific genes and a gene controlled by the enhancer of the MCK gene (*2*). In contrast, in CV1 cells, MyoD will activate a gene driven by the human cardiac actin gene enhancer (*33*) or by an oligomerized MyoD binding site (*26*), which suggests the presence of negative sequences in complicated enhanc-

ers of genes such as MCK. Failure to activate muscle-specific genes in HepG2 cells may be due to the fact that these cells do not express a positively acting, muscle-restricted factor that is required for MyoD activity (*57*); such a presumptive factor awaits isolation and its existence is not entirely consistent with the finding that MyoD can activate a gene containing oligomerized MyoD binding sites in these cells. Alternatively, HepG2 cells contain an activated *ras* oncogene (*58*), and, because this gene is already known to inhibit MyoD activity, it may not be surprising that HepG2 cells fail to activate muscle-specific genes; however, this explanation is also inconsistent with the fact that MyoD can activate a gene containing oligomerized MyoD binding sites. Clearly, the interactions between MyoD, oncogenes, antioncogenes, growth factors, and the cell background are very complex, and a clear picture has yet to emerge.

Factors that influence growth regulate MyoD, but MyoD itself can also inhibit cell growth (*59*). The first indications of this inhibition were in the initial studies demonstrating that LTR-MyoD could convert C3H $10T\frac{1}{2}$ cells to muscle (*1*); however, at that time it was unclear whether the inhibition of growth was caused by a direct effect on the cell cycle or as part of the program for muscle differentiation. Now two studies suggest that MyoD inhibits the cell cycle directly (*59*). When introduced into a variety of different types of virally transformed or malignant cells in culture, MyoD prevents entry into S phase. Because some of the cells expressing MyoD are known not to activate the myogenic program (such as CV1 cells and HeLa cells), and, because MyoD molecules with mutations in the basic region, which is required for myogenesis, still cause growth arrest, it would seem that MyoD can halt the cell cycle independent of its ability to induce myogenesis. Experiments with additional MyoD mutants show that the HLH domain is required for growth arrest, suggesting that growth arrest depends on complex formation with an HLH target (yet to be identified) but not upon DNA binding. In this respect, the inhibitory effect of MyoD on cell division is mechanistically similar to the inhibitory effect of Id or EMC on differentiation. The α factor in *Saccharomyces cerevisiae*, which both arrests cells at the G_1-S boundary (START) and induces a pathway of cell differentiation, shows many similarities to MyoD, which can also arrest cell division and induce differentiation (*60*).

Activation of *myoD* During Development

In primary fibroblasts, *myoD* expression is repressed by trans-acting factors. Somatic cell fusion experiments using micro-cell hybrids (*61*) show that, when human chromosome 11, which encodes *myoD*, is introduced from primary fibroblasts into permissive mouse C3H $10T\frac{1}{2}$ cells or a melanoma cell line, the human *myoD* gene on chromosome 11 is activated and causes these hybrid cells to differentiate into muscle. On the other hand, if additional human chromosomes (specifically 4 and 8) are transferred along with chromosome 11, activation of *myoD* and subsequent myogenesis does not occur, indicating that chromosomes 4 and 8 produce a trans-acting negative factor that inhibits expression of *myoD* on chromosome 11 and prevents myogenic activation of C3H $10T\frac{1}{2}$ cells. Loss of chromosome 4 from these hybrid lines containing chromosomes 4, 8, and 11 leads to the emergence of the muscle phenotype, suggesting that both 4 and 8 are required. Therefore, during development, *myoD* transcription may require, among other things, the removal of negative control. High levels of transcription might follow if autoactivation ensues.

We know little about how the *myoD* gene is activated during development. In four organisms—*Xenopus*, mice, *Drosophila*, and *C. elegans*—the bHLH domain of MyoD is remarkably conserved, and the MyoD protein is expressed only in skeletal muscle or their myogenic precursors, indicating that the MyoD function in determining muscle cell identity is a very ancient function. In contrast, although there is conservation of motifs for many of the *Drosophila* homeobox and Zn^{2+}-finger proteins involved in pattern formation, their patterns of expression in other organisms usually suggest that they perform different roles in different organisms. This suggests that the mechanisms for generating certain cell types such as muscle are ancient and highly conserved, but that each organism may have its own unique solution to the questions of how, where, and when a muscle cell is to be generated.

In *Xenopus*, mesoderm arises at the late blastula stage as a result of inductive interactions between vegetal pole blastomeres and overlying, animal pole–derived, equatorial blastomeres (*62*). The *myoD* gene is rapidly turned on when animal blastomeres are experimentally induced by juxtaposition with vegetal pole blastomeres (*63*). This interaction is now thought to be mediated by growth factors, as several such factors can induce muscle in isolated animal blastomeres. A major question is whether this induction of *myoD* by growth factors is a direct one or whether many gene activation events intervene.

Recently, MyoD has been cloned from *C. elegans* and an antibody to the protein produced (*64*). The antibody first stains cells at the 100-cell stage. These cells are lineage founder cells fated to give rise to body wall muscle descendants, some three to four cell generations later. Less than 3 kb of the upstream control region of *myoD* is needed for body wall muscle-specific expression, as revealed by fusing these sequences to a β-galactosidase expression vector. When assayed much earlier in development, embryos begin to express β-galactosidase before MyoD protein can be detected, at the 28-cell stage, but only in cells at least some of whose descendants give rise to muscle. Blastomeres at this stage that do not give rise to muscle descendants fail to express β-galactosidase, supporting the fidelity of the vector. These results suggest that the capacity for transcription of *myoD* in a myogenic lineage may precede the actual cell type–specific expression of the protein, and this capacity segregates into specific blastomeres as early as the 28-cell stage, a conclusion supported by both laser ablation experiments and cytochalasin experiments. The *myoD*–β-galactosidase fusions also suggest that precursor cells destined to give rise to descendants that become many different cell types can, in fact, express MyoD RNA. Presumably, subsequent cell-cell interactions determine which descendants will give rise to nerve and skin and turn off *myoD* transcription and which will maintain *myoD* expression and give rise to skeletal muscle cells. A similar conclusion has been reached about *Xenopus* animal pole cells in which MyoD mRNA is present before muscle induction, and the *myoD* expression is stabilized by either growth factors or interaction with vegetal cells (*65*).

Transcription in the absence of detectable translation has also been demonstrated with *achaete-scute*. RNA from *achaete-scute* is first detected in islands of about nine cells each in the presumptive neuroectoderm of the embryo (*66*). A single cell from each island then generates a sensory mother neuroblast (NB). This cell expresses the *achaete-scute* protein product, but the other eight cells do not. How this specific cell is chosen is not understood, although its effect on its neighbors is clear. The NB sends a lateral inhibitory signal to its neighbors to inhibit their differentiation into neuroblasts and to cause them to adopt the alternative pathway of becoming epidermis. This signal is mediated by the so-called neurogenic genes, including *notch*, *delta*, *enhancer of split*, and *shaggy*, which, when mutated, lead to extra bristles and neurons arising from each of the *achaete-scute* RNA-positive islands. Structural analysis of protein motifs encoded by several neurogenic genes suggests that these genes can define a pathway involved in cell adhesion, intercellular signaling, and coupling to nuclear events (*67*). The most downstream of these

genes is encoded by the *enhancer of split* locus which encodes bHLH proteins that may act as negative partners for *achaete-scute* products or Da in laterally inhibited presumptive epidermoblasts. Phenomena such as lateral inhibition are also observed in a number of organ systems in *C. elegans*, where the involved cells are termed an "equivalence group."

Many of the *achaete-scute* mutations were initially perplexing because the ability to produce neuroblasts was maintained by the organism as a whole, but specific neuroblasts were not formed in specific positions in the embryo. It now seems that the *achaete-scute* genes are controlled by multiple enhancers, each sensitive to the various positional coordinates previously established during embryogenesis (*68*). These coordinates may reflect expression of different combinations of segmentation, dorsal-ventral, and homeotic genes. Once *achaete-scute* is activated by positional information, it could then independently maintain its own expression if *achaete-scute* like *myoD*, can activate its own transcription. In contrast, muscle cell identity, at least in *Xenopus*, is determined by inductive interactions, whereas in *C. elegans* evidence suggests that at least some muscle lineages are determined by the segregation of maternal factors. It would seem that, although the mechanisms responsible for determination of specific cell types such as muscle or nerve are highly conserved, as evidenced by the apparent conservation of MyoD structure, function, and expression pattern, how *myoD* and *achaete-scute* are activated depends on the apparently individual strategies by which different organisms establish positional and temporal information.

Conclusion

The flow of cell type–specific information for myogenesis from the egg to the final muscle cell goes through a nodal point, defined by functions supplied by the MyoD family of myogenic regulatory proteins. It is possible that the *achaete-scute* complex of genes also acts as a nodal point, but for neurogenesis. An alternative organization could have also been imagined: combinations of less tissue-restricted genes could have encoded the *myoD* or *achaete-scute* functions. The nodal point need not have been an optimal solution but simply a solution that worked at some crucial time and in some context during evolution (*69*). Whether other cell lineages use this strategy and whether there is a profound insight provided by the use of such a strategy (such as the potential to make cell types mutually exclusive) await further analyses.

REFERENCES

1. R. L. Davis, H. Weintraub, A. B. Lassar, *Cell* **51**, 987 (1987).
2. H. Weintraub *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5434 (1989).
3. J. Choi *et al.*, *ibid.* **87**, 7988 (1990).
4. A.-Y. Lin, C. A. Dechesne, J. Eldridge, B. M. Paterson, *Genes Dev.* **3**, 986 (1989).
5. D. Sassoon *et al.*, *Nature* **341**, 303 (1989).
6. A. B. Lassar *et al.*, *Cell* **58**, 823 (1989).
7. R. Bischoff and H. Holtzer, *J. Cell Biol.* **44**, 134 (1970); F. Stockdale, K. Okazaki, M. Nameroff, H. Holtzer, *Science* **146**, 533 (1964); S. R. Dientsman and H. Holtzer, *Results Probl. Cell Differ.* **7**, 1 (1975); H. Holtzer *et al.*, *Q. Rev. Biophys.* **8**, 523 (1975).
8. M. C. O'Neil and F. E. Stockdale, *Dev. Biol.* **37**, 117 (1974).
9. S. J. Tapscott, A. B. Lassar, R. L. Davis, H. Weintraub, *Science* **245**, 532 (1989).
10. S. M. Taylor and P. A. Jones, *Cell* **17**, 771 (1979).
11. S. F. Konieczny and C. P. Emerson, Jr., *ibid.* **38**, 791 (1984).
12. P. A. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6117 (1990).
13. A. B. Lassar, B. M. Paterson, H. Weintraub, *Cell* **47**, 649 (1986); D. F. Pinney, S. H. Pearson-White, S. F. Konieczny, K. E. Latham, C. P. Emerson, Jr., *ibid.* **53**, 781 (1988).
14. J. B. Jaynes, J. E. Johnson, J. N. Buskin, C. L. Gartside, S. D. Hauschka, *Mol. Cell. Biol.* **6**, 2855 (1986).
15. R. A. Horlick and P. A. Benfield, *ibid.* **9**, 2396 (1989); E. A. Sternberg *et al.*, *ibid.* **8**, 2896 (1988); D. J. Bergsma, J. M. Grichnik, L. M. A. Gossett, R. J. Schwartz, *ibid.* **6**, 2462 (1986).
16. W. E. Wright, D. A. Sassoon, V. K. Lin, *Cell* **56**, 607 (1989).
17. T. Braun, G. Buschhausen-Denker, E. Bober, E. Tannich, H. H. Arnold, *EMBO J.* **8**, 701 (1989).
18. T. Braun, E. Bober, B. Winter, N. Rosenthal, H. H. Arnold, *ibid.* **9**, 821 (1990); J. H. Miner and B. Wold, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1089 (1990); S. J. Rhodes and S. F. Konieczny, *Genes Dev.* **3**, 2050 (1989).
19. M. J. Thayer *et al.*, *Cell* **58**, 241 (1989).
20. D. G. Edmondson and E. N. Olson, *Genes Dev.* **3**, 628 (1989); T. Braun *et al.*, *EMBO J.* **8**, 3617 (1989).
21. S. J. Tapscott *et al.*, *Science* **242**, 405 (1988).
22. C. Murre, P. S. McCaw, D. Baltimore, *Cell* **56**, 777 (1989).
23. R. L. Davis, P.-F. Cheng, A. B. Lassar, H. Weintraub, *ibid.* 60, 733 (1990).
24. R. Benezra, R. L. Davis, D. Lockshon, D. L. Turner, H. Weintraub, *ibid.* **61**, 49 (1990).
25. A. Voronova and B. Baltimore, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4722 (1990).
26. V. Dwarki, I. Verma, R. Davis, A. Lassar, H. Weintraub, unpublished observations.
27. H. Weintraub, R. Davis, D. Lockshon, A. Lassar, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5623 (1990).
28. C. Murre *et al.*, *Cell* **58**, 537 (1989).
29. X. H. Sun and D. Baltimore, *ibid.* **64**, 459 (1991).
30. P. Henthorn, M. Kiledjian, T. Kadesch, *Science* **247**, 467 (1990).
31. A. Lassar *et al.*, unpublished results.
32. M. Ptashne and A. Gann, *Nature* **346**, 329 (1990).
33. V. Sartorelli, K. Webster, L. Kedes, *Genes Dev.* **4**, 1811 (1990).
34. A. Ghysen and C. Dambly-Chaudiere, *Trends Genet.* **4**, 251 (1989).
35. M. Caudy *et al.*, *Cell* **55**, 1061 (1988).
36. R. Davis and H. Weintraub, unpublished observations.
37. D. Sakai *et al.*, *Genes Dev.* **2**, 1144 (1988); S. M. Hollenberg and R. Evans, *Cell* **55**, 899 (1988).
38. S. Stern, M. Tanaka, W. Herr, *Nature* **341**, 624 (1989).
39. K. Kim and L. Guarante, *ibid.* **341**, 200 (1989).
40. S. Tan and J. J. Richmond, *Cell* **62**, 367 (1990).
41. P. R. Mueller and B. Wold, *Science* **246**, 780 (1989).
42. R. Benezra and H. Weintraub, unpublished observations.
43. E. Olson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
44. H. M. Ellis, D. R. Spann, J. W. Posakony, *Cell* **61**, 27 (1990); J. Garrell and J. Modolell, *ibid.*, p. 39.
45. L. G. Alonso and A. Garcia-Bellido, *Roux's Arch. Dev. Biol.* **195**, 259 (1986); S. M. Parkhurst, D. Bopp, D. Ish-Horowicz, *Cell* **63**, 1179 (1990).
46. T. Cline, *Cell* **59**, 231 (1989).
47. F. Jimenez and J. Campos-Ortega, *Neuron* **5**, 81 (1990).
48. T. K. Blackwell and H. Weintraub, *Science* **250**, 1104 (1990).
49. N. D. Hopwood and J. B. Gurdon, *Nature* **347**, 197 (1990).
50. H. Holtzer, J. Brehl, G. Yeoh, R. Meganathan, A. Kaji, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4051 (1975).
51. A. B. Lassar, M. J. Thayer, R. W. Overell, H. Weintraub, *Cell* **58**, 659 (1989); S. F. Konieczny, B. L. Drobes, S. L. Menke, E. J. Taparowsky, *Oncogene* **4**, 473 (1989).
52. G. Fallone, F. Tato, S. Alema, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 426 (1985); S. A. Larocca, M. Gross, G. Fallone, S. Alema, F. Tato, *Cell* **58**, 123 (1989).
53. K. A. Webster, G. E. Muscat, L. Kedes, *Nature* **332**, 553 (1988).
54 T. Endo and B. Nadal-Ginard, *Cell* **49**, 515 (1987).
55. R. Cohen, M. Pacifici, N. Rubinstein, J. Biehl, H. Holtzer, *Nature* **266**, 538 (1977); A. A. Diugosz, S. J. Tapscott, H. Holtzer, *Cancer Res.* **43**, 2780 (1983).
56. H. J. Scrable *et al.*, *Genes Chromosomes Cancer* **1**, 23 (1989); H. J. Scrable, D. K. Johnson, E. M. Rinchik, W. K. Cavenee, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2182 (1990).
57. B. W. Schafer, B. T. Blakely, G. J. Darlington, H. M. Blau, *Nature* **344**, 454 (1990).
58. C. A. Richards, S. A. Short, S. S. Thorgersson, B. E. Huber, *Cancer Res.* **50**, 1521 (1990).
59. V. R. Sorrentino, R. Peperkok, R. L. Davis, W. Ansorge, L. Philipson, *Nature* **345**, 813 (1990); M. Crescenzi, T. P. Fleming, A. B. Lassar, H. Weintraub, S. A. Aaronson, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8442 (1990).
60. I. Herskowitz, *Nature* **342**, 749 (1989).
61. M. Thayer and H. Weintraub, *Cell* **63**, 23 (1990).
62. P. D. Nieuwkoop, *Wilhelm Roux's Arch. Entwicklungsmech. Org.* **162**, 341 (1969).
63. N. D. Hopwood, A. Pluck, J. B. Gurdon, *EMBO J.* **8**, 3409 (1989).
64. M. Krause, A. Fire, S. Harrison, J. Priess, H. Weintraub, *Cell* **63**, 907 (1990).
65. R. Rupp and H. Weintraub, unpublished data.
66. C. Cabrera, *Development (Camb.)* **109**, 733 (1990).
67. J. Campos-Ortega and E. Knust, *FEBS Lett.* 190, 1 (1990).
68. A. Ghysen and C. Dambly-Chaudiere, *Genes Dev.* **2**, 495 (1988).
69. F. Jacob, *The Possible and the Actual* (Pantheon, New York, 1982).