The EMBO Journal vol.8 no.3 pp.701 – 709, 1989

A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts

T.Braun, G.Buschhausen-Denker, E.Bober, E.Tannich and H.H.Arnold

University of Hamburg Medical School, Department of Toxicology, Grindelallee 117, D-2000 Hamburg 13, FRG

Communicated by M.Buckingham

We have isolated the cDNA encoding a novel human myogenic factor, Myf-5, by weak cross-hydridization to the mouse MyoD1 probe. Nucleotide sequence analysis and the identification of the corresponding gene indicate that human Myf-5 is a member of a small gene family which also contains the human homologue to MyoD1. Although structurally related to the mouse factor, the human Myf-5 constitutes a different protein which nevertheless is capable of inducing the myogenic phenotype in embryonic C3H mouse 10T1/2 'fibroblasts'. The existence of more than one MyoD1-like protein in human skeletal muscle is further suggested by the detection of several similar but distinct cDNA clones. The phenotypic conversion of 10T1/2 cells by the human factor is recognized by the capacity of the cells to form multinucleated syncytia and synthesize sarcomeric myosin heavy chains. Moreover, transient expression of Myf-5 in 10T1/2 cells leads to the activation of a co-transfected muscle-specific CAT reporter gene which by itself is transcriptionally silent in the non-muscle cell background. The deduced amino acid sequence of clone Myf-5 reveals a region which is highly similar to myc proteins and the developmental factors from *Drosophila* encoded by the *achaete scute* locus and the *twist* gene. The myc homology region and a preceding cluster of basic amino acids are located in a larger sequence domain with strong similarity to the mouse myogenic factor MyoD1. Two additional short segments with high serine and threonine content are conserved between the two proteins. Considering the similar activities of human Myf-5 and mouse MyoD1 in the biological assay of myogenic conversion and the segmental conservation of distinct sequence domains, it is suggested that the three conserved sequence regions (including the myc homology) play a critical role in the myogenic programming of mesodermal cells.

Key words: human Myf-5/muscle determination/Myc-homology/transactivation

Introduction

The original observation that C3H 10T1/2 mouse embryonic 'fibroblasts' can be converted into three mesenchymal cell types, i.e. myogenic, adipogenic and chondrogenic cells, by a brief treatment with 5-azacytidine (Taylor and Jones, 1979) stimulated research addressing the molecular mechanisms of the cellular commitment. It was speculated that the incorporation of 5-azacytidine into DNA causes demethylation of regulatory genetic loci which are involved in the determination of the observed phenotypes (Jones and Taylor, 1980). The stable and heritable effect of the drug and the high frequency of myogenic conversion led to the assumption that one or only few closely linked genes would be sufficient to convert 10T1/2 cells into myoblasts (Konieczny and Emerson, 1984). It was indeed demonstrated that transfection of DNA from 5-azacytidine-derived 10T1/2 myoblasts, but not DNA from the original C3H 10T1/2 fibroblasts, causes muscle cell formation at a frequency consistent with the transmission of a single activated gene (Konieczny *et al.*, 1985; Lassar *et al.*, 1986). Subsequently, it was shown that the cDNA clone MyoD1, which was isolated from a 10T1/2-cell-derived myoblast line by differential cDNA hybridization (Davis *et al.*, 1987), is capable of converting several fibroblast and adipoblast cell lines into determined myoblasts upon expression of the transfected DNA. More recently, it was reported that transfection of human genomic DNA cloned into a cosmid vector also induces myogenesis in 10T1/2 cells (Pinney *et al.*, 1988). This putative myogenic determination gene, however, did not correspond to the mouse MyoD1 gene and seemed to be expressed in myoblasts prior to MyoD1.

To study the structure and function of a human factor which is involved in muscle development, we isolated several cDNA clones from fetal human skeletal muscle by their weak cross-hybridization to the mouse MyoD1 probe. The complete clone Myf-5, when expressed under the control of a viral promoter, converts 10T1/2 cells to myoblasts at high frequency. The deduced protein sequence of Myf-5 exhibits an interesting segmental conservation to mouse MyoD1, particularly in a domain that contains a region of basic amino acids and a strong homology to cellular and viral myc proteins. The highly conserved structure of three protein domains in mouse MyoD1 and human Myf-5 and the same biological activity in the 10T1/2 myogenic conversion assay strongly suggest that the homology regions are involved in the myogenic programming.

Results

Isolation of human cDNA clones related to mouse MyoD1

Using the mouse MyoD1 probe, we screened a human cDNA library from fetal skeletal muscle under hybridization conditions of reduced stringency (see Materials and methods). Approximately 30 cDNA clones were isolated out of 5×10^6 recombinant phages. They were grouped into three categories according to their insert size and slightly different capacity to cross-hybridize to MyoD1 (Figure 1a). Three clones, one of each group, were selected for comparative restriction analysis. As shown in Figure 1e, the three clones (called Myf-3, Myf-4 and Myf-5, although obviously not all full length) shared no common restriction sites except for one *Pst*I site conserved in Myf-4 and Myf-5,

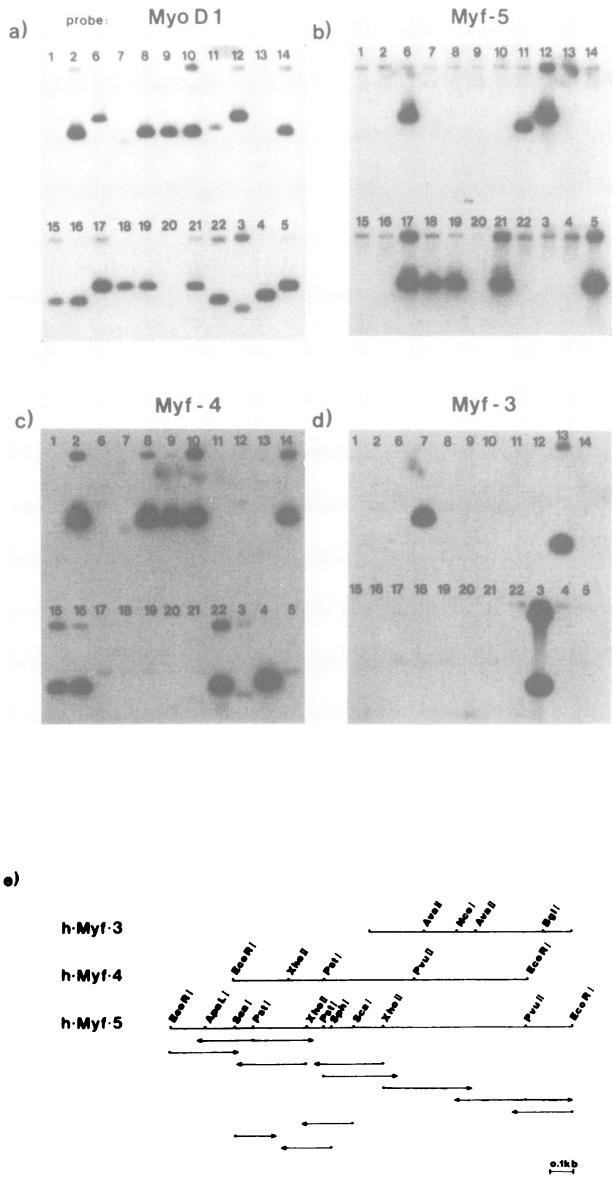

Fig. 1. Human cDNA clones isolated with the mouse MyoD1 probe from a gt11 cDNA library of human skeletal muscle. Isolated gt11 recombinants were digested with *Eco*RI to delineate the cDNA inserts. The DNA blots were hybridized successively to: (**a**) ^{32}P-labelled mouse MyoD1 probe; (**b**) a Myf5-specific fragment (*Xho*II/*Eco*RI); (**c**) a Myf4-specific probe (*Pvu*II/*Eco*RI); and (**d**) the total Myf3-specific insert. Clones Myf-3, Myf-4 and Myf-5 were selected as representatives of the three types of MyoD1-related clones. The individual length and partial restriction maps of the three clones, as well as the sequencing strategy for clone Myf-5 are illustrated (**e**). The arrows indicate the direction and extent of individual sequencing runs.

indicating their very limited structural relatedness. This fact was also demonstrated by differential hybridizations using specific parts of the three clones as hybridization probes (Figure 1b–d). The nucleotide sequence of the most complete cDNA clone Myf-5 was determined. The sequence of 1430 nucleotides (nt) together with the deduced putative amino acid sequence encoded by the single open reading frame (ORF) in both orientations is shown in Figure 2. According to this analysis, Myf-5 codes for a protein of 255 amino acids which is 63 amino acids shorter than mouse MyoD1. Although the largest ORF begins with

Myf-5

```
        10         20         30         40         50         60
CCTCTCGCTGCCGTCCAGGTGCACCGCCTGCCTCTCAGCAGGATGGACGTGATGGATGGC
                                                MetAspValMetAspGly

        70         80         90        100        110        120
TGCCAGTTCTCACCTTCTGAGTACTTCTACGACGGCTCCTGCATACCGTCCCCCGAGGGT
CysGlnPheSerProSerGluTyrPheTyrAspGlySerCysIleProSerProGluGly

       130        140        150        160        170        180
GAATTTGGGGACGAGTTTGTGCCGCGAGTGGCTGCCTTCGGAGCGCACAAAGCAGAGCTG
GluPheGlyAspGluPheValProArgValAlaAlaPheGlyAlaHisLysAlaGluLeu

       190        200        210        220        230        240
CAGGGCTCAGATGAGGACGAGCACGTGCGAGCGCCTACCGGCCACCACCAGGCTGGTCAC
GlnGlySerAspGluAspGluHisValArgAlaProThrGlyHisHisGlnAlaGlyHis

       250        260        270        280        290        300
TGCCTCATGTGGGCCTGCAAAGCCTGCAAGAGGAAGTCCACCACCATGGATCGGCGGAAG
CysLeuMetTrpAlaCysLysAlaCysLysArgLysSerThrThrMetAspArgArgLys

       310        320        330        340        350        360
GCAGCCACTATGCGCGAGCGGAGGCGCCTGAAGAAGGTCAACCAGGCTTTCGAAACCCTC
AlaAlaThrMetArgGluArgArgArgLeuLysLysValAsnGlnAlaPheGluThrLeu

       370        380        390        400        410        420
AAGAGGTGTACCACGACCAACCCCAACCAGAGGCTGCCCAAGGTGGAGATCCTCAGGAAT
LysArgCysThrThrThrAsnProAsnGlnArgLeuProLysValGluIleLeuArgAsn

       430        440        450        460        470        480
GCCATCCGCTACATCGAGAGCCTGCAGGAGTTGCTGAGAGAGCAGGTGGAGAACTACTAT
AlaIleArgTyrIleGluSerLeuGlnGluLeuLeuArgGluGlnValGluAsnTyrTyr

       490        500        510        520        530        540
AGCCTGCCGGGACAGAGCTGCTCGGAGCCCACCAGCCCCACCTCCAACTGCTCTGATGGC
SerLeuProGlyGlnSerCysSerGluProThrSerProThrSerAsnCysSerAspGly

       550        560        570        580        590        600
ATGCCCGAATGTAACAGTCCTGTCTGGTCCAGAAAGAGCAGTACTTTTGACAGCATCTAC
MetProGluCysAsnSerProValTrpSerArgLysSerSerThrPheAspSerIleTyr

       610        620        630        640        650        660
TGTCCTGATGTATCAAATGTATATGCCACAGATAAAAACTCCTTATCCAGCTTGGATTGC
CysProAspValSerAsnValTyrAlaThrAspLysAsnSerLeuSerSerLeuAspCys

       670        680        690        700        710        720
TTATCCAACATAGTGGACCGGATCACCTCCTCAGAGCAACCTGGGTTGCCTCTCCAGGAT
LeuSerAsnIleValAspArgIleThrSerSerGluGlnProGlyLeuProLeuGlnAsp

       730        740        750        760        770        780
CTGGCTTCTCTCTCTCCAGTTGCCAGCACCGATTCACAGCCTCGAACTCCAGGGGCTTCT
LeuAlaSerLeuSerProValAlaSerThrAspSerGlnProArgThrProGlyAlaSer

       790        800        810        820        830        840
AGTTCCAGGCTTATCTATCATGTGCTATGAACTAATTTTCTGGTCTATATGACTTCTTCC
SerSerArgLeuIleTyrHisValLeuEnd

       850        860        870        880        890        900
AGGAGGGCCTAATACACAGGACGAAGAAGGCTTCAAAAAGTCCCAAACCAAGACAACATG

       910        920        930        940        950        960
TACATAAAGATTTCTTTTCAGTTGTAAATTTGTAAAGATTACCTTGCCACTTTATAAGAA

       970        980        990       1000       1010       1020
AGTGTATTTAACTAAAAAGTCATCATTGCAAATAATACTTTCTTCTTCTTTATTATTCTT

      1030       1040       1050       1060       1070       1080
TGCTTAGATATTAATACATAGTTCCAGTAATACTATTTCTGATAGGGGGCCATTGATTGA

      1090       1100       1110       1120       1130       1140
GGGTAGCTTGTTCGAATGCTTAACTTATATATACATATATATATTATAAATATTGCTC

      1150       1160       1170       1180       1190       1200
ATCAAAATGTCTCTGGTGTTTAGAGCTTTATTTTTTTCTTTAAAACATTAAAACAGCTGA

      1210       1220       1230       1240       1250       1260
GAATCAGTTAAATGGAATTTTAAATATATTTAACTATTTCTTTTCTCTTTAATCCTTTAG

      1270       1280       1290       1300       1310       1320
TTATATTGTATTAAATAAAAATATAATACTGCCTAATGTATATATTTTGATCTTTTCTTG

      1330       1340       1350       1360       1370       1380
TAAGAAATGTATCTTTTAAATGTAAGCACAAAATAGTACTTTGTGGATCATTTCAAGATA

      1390       1400       1410       1420       1430
TAAGAAATTTTGGAAATTCCACCATAAATAAAATTTTTTACTACAAGAAAAA
```

Fig. 2. The nucleotide sequence and the deduced amino acid sequence of the cDNA clone Myf-5.

CAGGATGG, a reasonably good match of the vertebrate initiation consensus sequence (Cavener, 1987), we cannot rule out the possibility that the first AUG is not the trans-

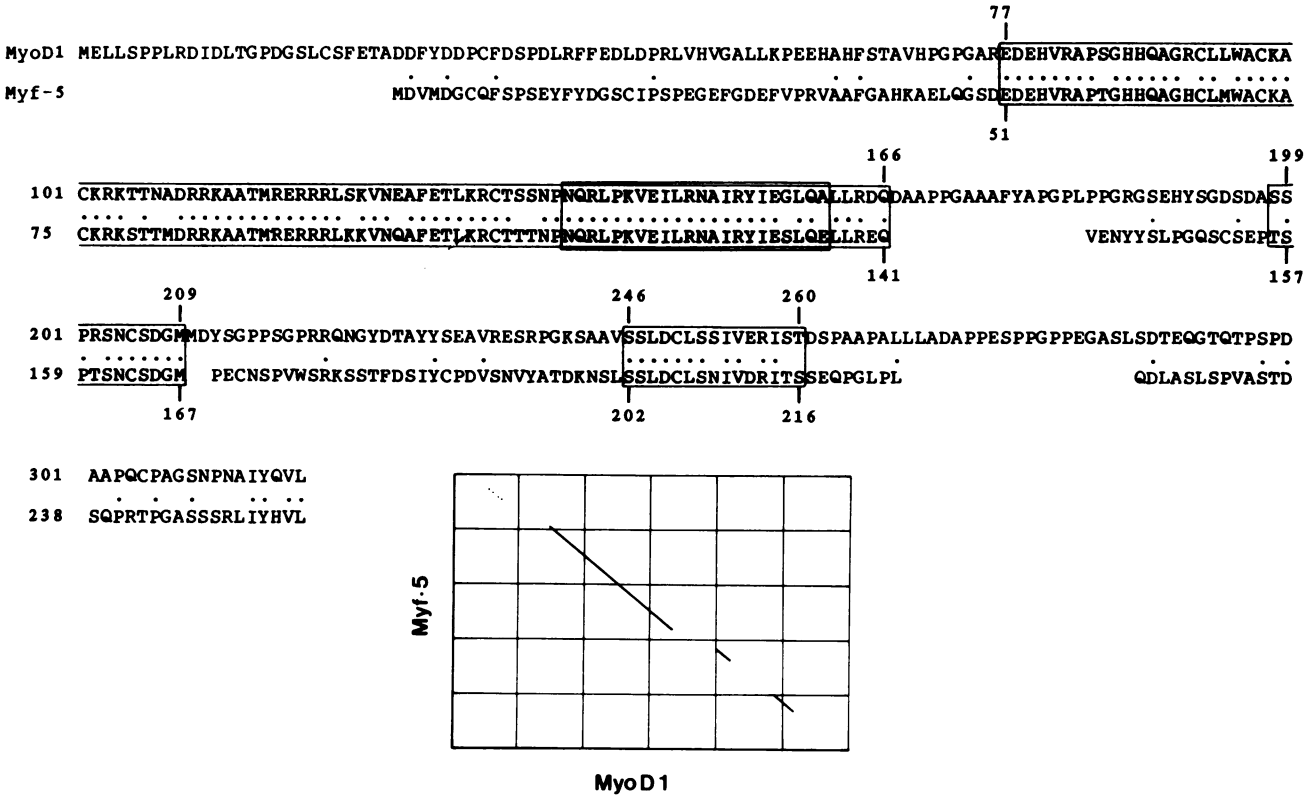

Fig. 3. Protein sequence comparison of mouse MyoD1 and human Myf-5. Segments of homologies shared between MyoD1 (aa 77–166, 199–209, 246–260) and Myf-5 (aa 51–141, 157–167, 202–216) are framed and shaded. Gaps in the Myf-5 sequence were introduced for optimal alignment of the homologous regions. Pairs of identical amino acids are indicated by dots. The myc homology region is indicated by the double-lined area. The insert shows the homology plot for the nucleotide sequence of mouse MyoD1 (horizontal axis) and human Myf-5 (vertical axis). Homology was scored by a dot if 6 out of 8 bases were identical.

lational start site and the clone is not complete at the 5′ end. The 3′ non-coding sequence following the TGA stop codon is 620 nt long, similar to the 3′ untranslated region of the mouse MyoD1 cDNA. Interestingly, this relatively long sequence contains 74% A + T comprising several extended runs of A and T which also occur in numerous lymphokine, cytokine and proto-oncogene mRNAs leading to specific RNA destabilization (Shaw and Kamen, 1986). The 'destabilizer' sequence motif of a single adenosine nucleotide, a polyuridine tract of three or more Us followed by an adenosine, exists three times in the 3′ non-coding sequence of Myf-5. In addition, several imperfect repeats of this consensus sequence are present in the Myf-5 3′ end sequence.

In a Northern blot analysis a mRNA species of ~1.7 kb is found in fetal and adult skeletal muscle tissues and also in smooth muscle from uterus. No Myf-5 mRNA was detected in fetal heart, non-muscle tissues and HeLa cells (data not shown).

The human Myf-5 protein contains a myc homology region and three sequence domains which are conserved to mouse MyoD1

When the protein sequence of human Myf-5 was deduced and compared to the mouse MyoD1 structure, three segments of amino acid sequence conservation were detected (Figure 3). Both proteins start at the N termini with entirely different sequences of 50 and 77 amino acids in Myf-5 and MyoD1 respectively. This sequence is followed by a block of 91 amino acids which are highly homologous in both proteins. The sequence conservation was calculated to be 86% including the conservative amino acid changes (Glu–Asp, Arg–His, Thr–Ser). Following a stretch of 31 and 15 non-conserved amino acids in MyoD1 and Myf-5 respectively, a second block of 11 amino acids exhibits 82% sequence similarity or 90% of sequence conservation if functionally equivalent substitutions are not counted (Thr–Ser). A third conserved sequence segment is located ~35 amino acids downstream and consists of 15 amino acids of 73% homologous sequence (93% without the conservative substitutions). This segmental sequence conservation between the mouse MyoD1 and the human Myf-5 could also be demonstrated by a homology plot of the nucleotide sequences (Figure 3). The most striking feature of the largest (91 amino acid) conserved sequence domain is a short segment (22 amino acid) which exhibits strong similarity to a conserved domain of myc proteins from various sources (Figure 4A). The same sequence block is also conserved in the *achaete scute* genes of *Drosophila* which are involved in neuronal and sex development (Alonso and Cabrera, 1988) and in the *Drosophila twist* gene product (Thisse *et al.*, 1988) (see Figure 4B). For the human c-myc protein it has been shown that this conserved region is of functional importance for the transforming capacity *in vitro* (Stone *et al.*, 1987). The myc homology region is preceded by a highly significant basic region which is also present in the *Drosophila achaete scute* gene products (shown for T4) as well as in *twist* gene product (Figure 4B). Both the basic region and the myc

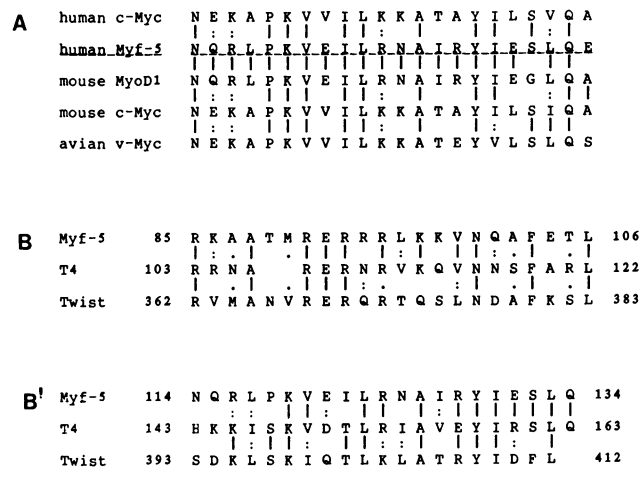

Fig. 4. Myf-5 sequence domains shared by myc proteins, achaete scute protein and the *twist* gene product of *Drosophila*. (**A**) Amino acids with sequence homology between various myc-proteins and Myf-5 are shown. Identical amino acids are indicated by dashes, conservative substitutions are shown by double dots. (**B**) Sequence comparison of the basic region between Myf-5, the achaete scute protein T4 and the *twist* gene product. (**B'**) Sequence comparison of the myc-homology region in Myf-5, the achaete scute protein T4 and the *twist* gene product. The coordinates indicate the amino acids in the respective proteins.

homology region have already been shown to exist in mouse MyoD1 (Davis *et al.*, 1987). In addition, a putative metal-binding domain of the 'zinc-finger' class (Berg, 1986) was postulated for MyoD1 between residues 62 and 101 (Davis *et al.*, 1987). The N-terminal part of this sequence (62–77) is not conserved in human Myf-5, which particularly lacks two histidine residues and therefore does not conform to the postulated sequence requirements for finger structures. The two smaller segments of sequence conservation located towards the C-terminal region show a remarkably high content of serine and threonine residues (5 of 11 and 6 of 15 respectively) which are noticeably frequent in the entire C-terminal half of the molecule. The clustered appearance of these amino acids could indicate potential sites of protein phosphorylation.

Constitutive expression of human Myf-5 converts mouse 10T1/2 fibroblasts into myogenic cells

Human Myf-5 cDNA was cloned into a vector (pEMSV scribe α) containing the LTR region of the Molony sarcoma virus and the SV40 polyadenylation signal for efficient transcription in eukaryotic cells. Constructs containing the cDNA in sense (pEMSV-Myf-5) or antisense orientation (pEMSV-con) were co-transfected together with pSV2-neo plasmid carrying the selectable marker gene which confers G418 resistance. A molar ratio of non-selectable to selectable plasmid DNA of 30:1 was chosen to ensure efficient co-transfer of Myf-5 DNA. Selection for stably transfected cell clones was started 24 h after transfection. The myogenic potential in G418-resistant colonies was tested by the appearance of fused, multinucleated myotubes following the shift to fusion-supporting medium (see Materials and methods). Cells of differentiating colonies at early stages were recloned for the isolation and clonal expansion of Myf-5-derived myoblasts. For the documentation of differentiated colonies the cells were fixed and stained with Giemsa. As shown in Table I, >50% of G418-resistant

Table I. Efficiency of myogenic conversion by Myf-5

Transfected cells	Transfected DNA vectors (myogenic colonies[a] per neo-resistant colonies)		
	pSV2-neo[b]	pSV2-neo[c] pEMSV-Myf5	pSV2-neo[d] pEMSV-con
C3H 10T1/2 mouse fibroblast	0/80	42/78	0/54
MCA CL15 MC-transformed mouse fibroblast	0/95	54/83	0/38

[a]Myogenic colonies were scored by the morphological appearance of multinucleated myotubes 4–5 days after medium shift.
Standard transfections on T-25 flasks were performed with:
[b]1 μg pSV2-neo plus 30 μg pUC-18 plasmid as carrier;
[c]1 μg pSV2-neo plus 30 μg pEMSV-Myf5;
[d]1 μg pSV2-neo plus 30 μg pEMSV-con.

colonies which had been co-transfected with Myf-5 in the sense orientation were capable of differentiating into myotubes. We have never observed myogenic differentiation in colonies which had been obtained from transfections with either pSV2-neo plasmid alone, or pSV2-neo plus pEMSV-con plasmid (Myf-5 cDNA in the antisense orientation). The myotubes formed in Myf-5-expressing colonies were characterized by large numbers of closely packed nuclei in syncytia of bag- or tube-like structures (Figure 5). These muscle cells were morphologically indistinguishable from 10T1/2 myotubes derived from short exposure to 5-aza-cytidine or by transfection with the mouse MyoD1 expression vector, pEMSV-scribe MyoD. To confirm the muscle nature of the converted cells, colonies were fixed and analysed for myosin heavy chain expression with specific antibodies. As shown in Figure 5, differentiated muscle cells stained positively with the anti-myosin heavy chain antibody, indicating that not only fusion but also the appropriate expression of a contractile protein was inducible in Myf-5-transfected 10T1/2 cells, but not in the untransfected 10T1/2 recipient cells (control not shown). It is interesting to note that myogenic conversion and expression of myosin was obtained not only in the established fibroblast line C3H 10T1/2, but also in the methylcholanthrene-transformed derivative of 10T1/2 cells, MCA-CL15 (Figure 5B). As shown in Table I, based on morphological criteria the tumorigenic cell line MCA-CL15 (Suhuin *et al.*, 1986) was converted to myogenic colonies at a similar rate as the non-tumorigenic C3H 10T1/2 cells. This seems to indicate that Myf-5 exerts a dominant determination effect over the malignant phenotype of these chemically transformed tumour cells.

It was demonstrated by Davis *et al.* (1987) that the mouse MyoD1 gene becomes activated in myogenic 10T1/2-derived cells and hence can be regarded as an early marker for muscle development. To investigate how the expression of the endogenous MyoD1 gene would be affected by the expression of the human Myf-5 cDNA in 10T1/2 and MCA C115 cells, RNA was analysed on a Northern blot. As shown in Figure 6, MyoD1 specific mRNA was abundantly expressed in myogenic colonies derived from 10T1/2 cells but undetectable in the parental 10T1/2 cells. In contrast, myogenic colonies derived from MCA-CL15 cells expressed no or very little MyoD1 mRNA, although they exhibit the comparable capacity to differentiate to muscle cells. This observation cannot be attributed to there being little or no

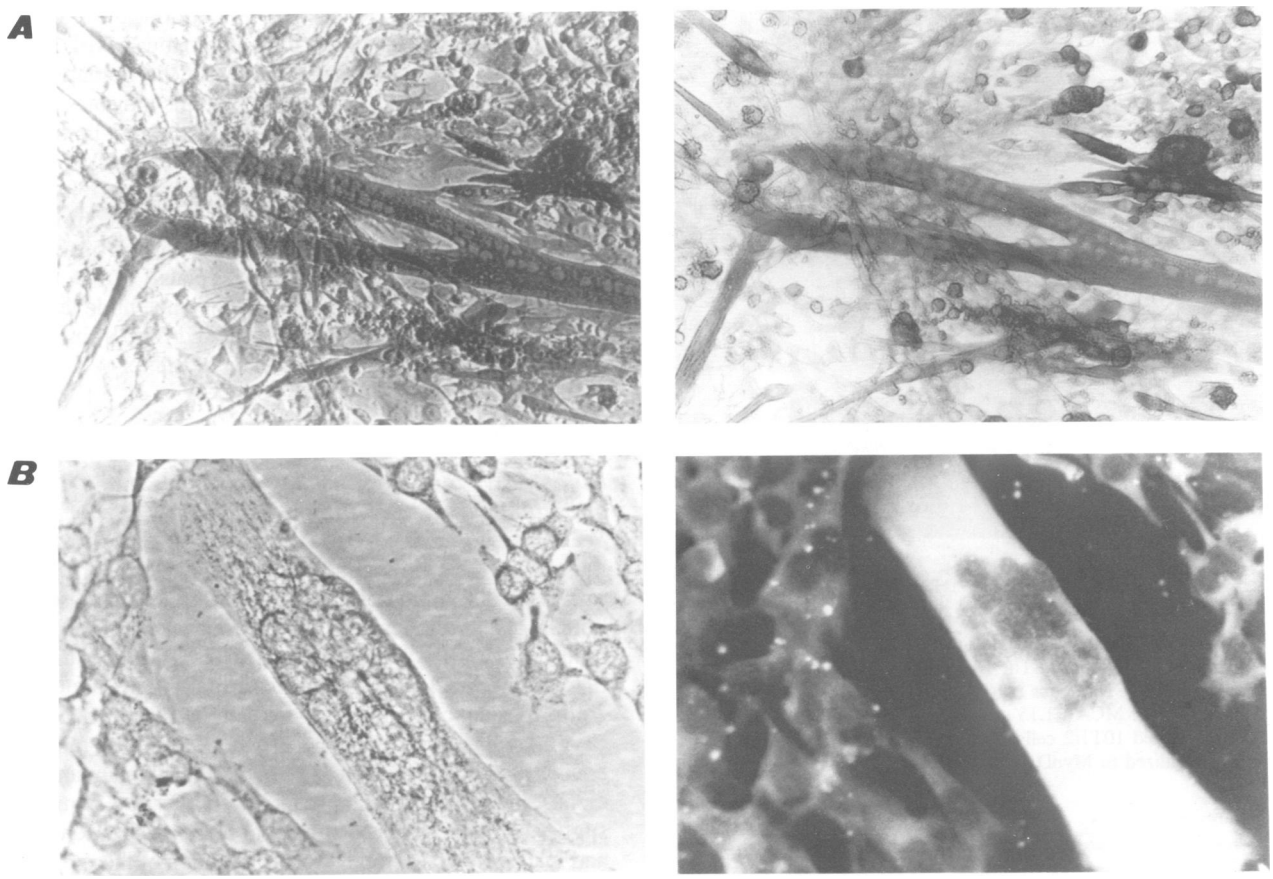

Fig. 5. Phase contrast micrographs (**left**) and immunostaining with antimyosin heavy-chain antibody (**right**) of 10T1/2 fibroblasts (**A**) or MLCA CL15 cells (**B**) converted with human pEMSV-Myf5. Immunopositive cells visualized with horseradish peroxidase-coupled second antibody (**A**) appear dark in the photograph, those stained with fluorescence-coupled antibody (**B**) appear light.

expression of Myf-5 in those cells, since Myf-5-specific mRNA was readily detectable in transfected MCA-CL15. Taken together these observations might indicate that the mouse MyoD1 gene can be activated as a result of the Myf-5 expression, but that this activation is not a prerequisite for myogenic conversion.

Transient expression of Myf-5 activates muscle-specific promoters in non-muscle cells

The myosin conversion assay of 10T1/2 mouse fibroblasts, as described above, involves a complex pattern of morphological changes and alterations of gene activities. This 'muscle development *in vitro*' seems to require not only the stable expression of the transfected Myf-5 cDNA but in addition the withdrawal of serum from the culture medium in order to promote terminal differentiation. To test the potential transactivating capacity of the human Myf-5 in a short-term assay, a suitable transient transfection system had to be established. We have previously demonstrated that the promoter of the chicken regulatory myosin light-chain gene (MLC2-A) is transcribed only in differentiated primary myocytes but not in established myogenic cell lines, proliferating myoblasts or fibroblasts (Arnold *et al.*, 1988; Braun *et al.*, 1988a,b). Accordingly, when the LC Δ-64, a chloramphenicol acetyl transferase (CAT) reporter construct, is transfected into mouse 10T1/2 fibroblasts no CAT activity was obtained (Figure 7). The *trans*-activating competence of the Myf-5 gene product was demonstrated when the expression vehicle pEMSV-Myf5 was cotransfected with the muscle-specific MLC2−CAT reporter construct. As shown in Figure 7, transient expression of human Myf-5 results in the concomitant activation of the muscle-specific chicken MLC2 promoter in an established mouse non-muscle cell line. In contrast, the expression of a constitutive control gene, such as the cytoplasmic β-actin promoter in the P1−CAT construct, remains virtually unaffected. Stimulation of promoter activity was also observed for the chicken cardiac actin promoter (Quitschke *et al.*, 1987) which is not completely repressed in fibroblasts but nevertheless responds to Myf-5 with increased transcriptional activity. The activation of the MLC2 promoter shows a concentration dependence of the co-transfected Myf-5 expression vehicle. The observed CAT activity directed by constant amounts of reporter plasmid (LCΔ-64) increases linearly with the concentration of pEMSV-Myf5 plasmid (data not shown). It should be mentioned that in the described experiments CAT activity was determined 4 days after DNA transfection, in growth-promoting medium (10% serum), when no morphological change (fusion) was detectable in these cultures.

Myf-5 is a member of a small group of genes sharing moderately related common sequences

The isolation of three different types of cDNA clones from human skeletal muscle (see Figure 1) by cross-hybridization to mouse MyoD1 was suggestive of either one gene generating different but related Myf-mRNAs by alternative splicing pathways or the existence of more than one gene

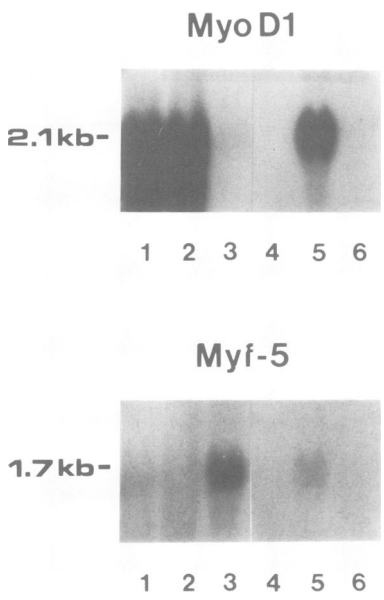

Fig. 6. Northern blot analysis using mouse MyoD1 and human Myf-5 specific hybridization probes. The lanes contain RNA from undifferentiated C2C12 mouse myoblasts (**1**), C2C12 myotubes (**2**), Myf-5 transformed MCA CL15 cells (**3**), MCA CL15 (**4**), Myf-5-transformed 10T1/2 cells (**5**) and 10T1/2 fibroblasts (**6**). The blot was hybridized to MyoD1 cDNA (**A**) or Myf-5 cDNA (**B**).

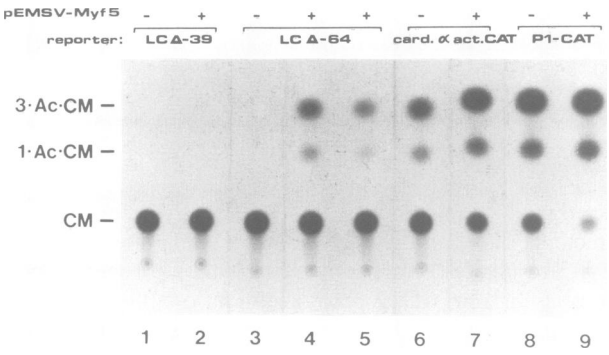

Fig. 7. Activation of muscle promoters in mouse 10T1/2 fibroblasts transiently transfected with pEMSV-Myf5. 10 μg of reporter constructs LCΔ-39 (no promoter control), LCΔ-64 (chicken MLC 2A promoter), cardiac α-actin CAT (chicken cardiac actin promoter), and P1-CAT (chicken cytoplasmic β-actin promoter) were co-transfected with 10 μg of pUC8 plasmid DNA (−) or 10 μg of pEMSV-Myf5 DNA (+) into 10T1/2 fibroblasts and CAT activity was determined 4 days later as described in Materials and methods. In lane 5, 5 μg of LCΔ-64 reporter was used for co-transfection.

giving rise to several Myf transcripts. To test these alternatives, human genomic DNA and mouse DNA for comparison were digested with several restriction endonucleases and probed with total Myf-5 cDNA. As shown in Figure 8, under moderately stringent conditions one or only a few prominent hybridization signals accompanied by few much less intense bands appeared in all lanes. When probes specific for Myf-4 or Myf-3 were used, distinct sets of signals could be detected. There was no overlapping banding pattern of the genomic blots among the three different probes but rather the indication that at least three different Myf genes exist in man. Direct cross-hybridization of human DNA with the mouse MyoD1 probe, however, leads to the detection of a banding pattern which overlaps the pattern generated by Myf-3. In addition, the fragments detected in mouse DNA with the homologous MyoD1 probe clearly differ from those detected with the human probes for Myf-4 and Myf-5. Taken together, this indicates that the gene encoding Myf-3 is likely to be the human homologue for mouse MyoD1. This interpretation is also supported by nucleotide sequence analysis of the partial cDNA clone Myf-3 and its greater homology to MyoD1 as compared to Myf-5 (unpublished results). Within the regions of sequence similarities shared among the three isolated cDNA clones, only incomplete homology exists at the nucleotide level and therefore the transcripts cannot be derived from a single gene by alternative splicing (unpublished results).

Discussion

Human Myf-5 is functionally equivalent and structurally related to mouse MyoD1

We have isolated several human cDNAs from skeletal muscle by their limited structural similarity to the mouse myogenic determination factor, MyoD1. The longest of the human cDNA clones, Myf-5, has been shown to convert mouse C3H 10T1/2 fibroblasts to myoblasts which differentiate to multinucleated myotubes virtually indistinguishable from muscle cells that have been derived with mouse MyoD1. The observation of the apparent functional equivalence of both the mouse MyoD1 and the human Myf-5 clones in the genetic approach of phenotype conversion suggested that conserved structural features between mouse and man would be essential for the function of these proteins. The interesting result of a detailed structural comparison was that only certain segments of the two proteins were highly conserved, whereas the rest of the structures was neither similar in length nor in amino acid sequence. In the N-terminal half of the molecule a large block of ~90 homologous amino acids exists. The other two blocks of conserved sequences in Myf-5 are located in the C-terminal half and are only 11 and 16 amino acids long and exceptionally rich in serine and threonine, representing potential sites for protein phosphorylation. From these observed structural similarities in three distinct domains of MyoD1 and Myf-5 and from the equal capacity of both proteins to convert mouse 10T1/2 fibroblasts to myogenic cells, we conclude that the conserved structures are most likely to be critical for the biological function.

Myf-5 contains regions of homology to myc proteins and developmental factors of Drosophila; it activates muscle genes 'in trans'

The largest stretch of conserved sequence in Myf-5 contains a pronounced basic cluster, followed by 22 amino acids with a high degree of sequence similarity to myc proteins of various origins. The conserved myc-homology region was also found in the human Myf-4 clone which is otherwise very little related to Myf-5 (preliminary results from this laboratory). The same block of conserved myc sequence is also present in the *achaete scute* genes and the *twist* gene of *Drosophila* which are involved in neurogenic development (Cabrera *et al.*, 1987; Alonso and Cabrera, 1988; Villares and Cabrera, 1987) and the establishment of germ layers (Thisse *et al.*, 1988) respectively. In the myc protein, this region is important for two-step transformation (Stone *et al.*, 1987). It has also been demonstrated that myc proteins are capable of activating transcription in a yeast hybrid protein

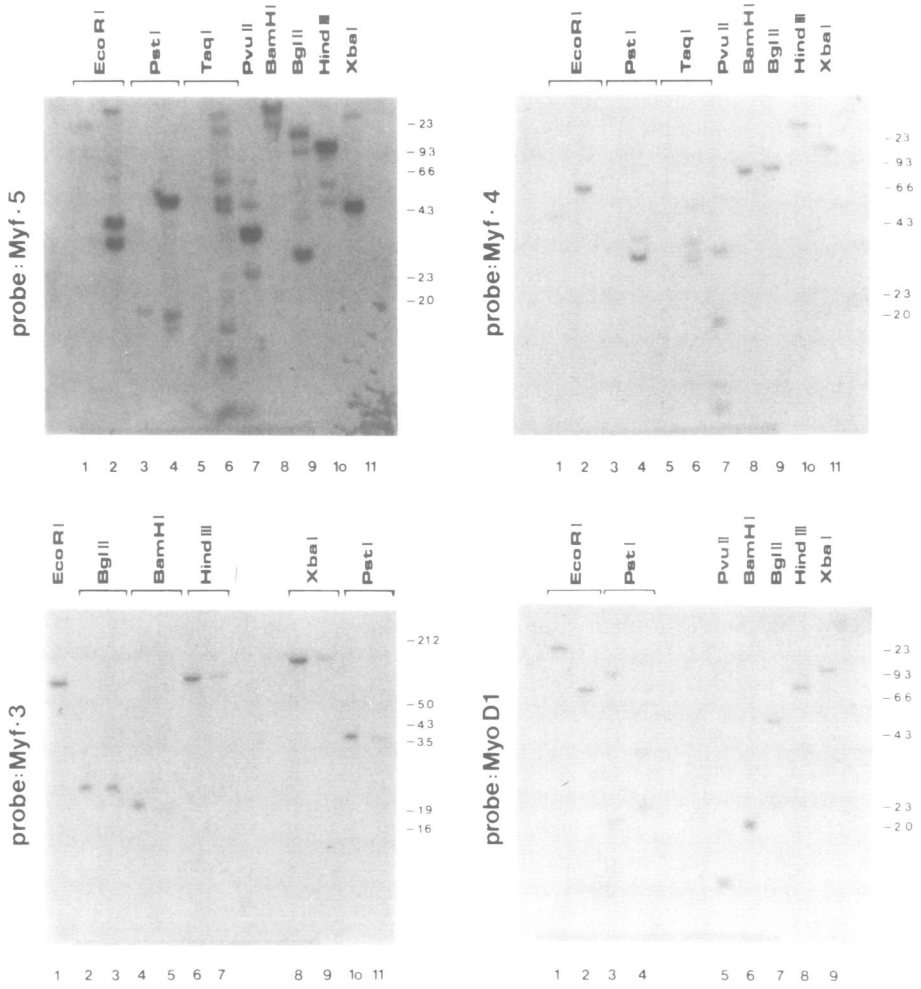

Fig. 8. Hybridization of mouse and human genomic DNA with probes isolated from human cDNA clones Myf-5, Myf-4, Myf-3 and mouse MyoD1. DNA was digested with the indicated restriction endonucleases and analysed by the Southern blot technique. Autoradiographs shown for the probes Myf-5, Myf-4 and MyoD1 are successive rehybridizations of the same blot after removal of prior hybridization signals. **Lanes 1, 3** and **5** of blots 1, 2 and 4 contain mouse DNA for comparison. All other lanes contain human DNA. With mouse MyoD1 and Myf-3 no *Taq*I-digested fragments could be detected, therefore these lanes have been omitted from the numbering. The blot hybridized with Myf-3 contains only human DNA of two concentrations (5 and 10 μg for each restriction enzyme) in adjacent lanes, except for *Eco*RI where only 10 μg of DNA was used. *Hin*dIII digested λ DNA was used as marker.

assay (Lech *et al.*, 1988). Furthermore, myc proteins were reported to cause '*trans*-activation' by stimulating transcription of transiently expressed transfected reporter genes (Kingston *et al.*, 1984; Kaddurah-Daouth *et al.*, 1987). Based on the myc sequence homology, it seemed therefore possible that Myf-5 protein is also a transcription factor which acts in a similar manner to the myc proteins. This idea is supported by the result that Myf-5, when transiently expressed in non-muscle cells, stimulates transcription of muscle-specific promoters. By the myc analogy, one could assume that the basic domain is involved in DNA binding. Since no direct interaction of Myf-5 protein and the transcriptionally stimulated promoters has been shown, however, an indirect mechanism of muscle gene activation by Myf-5 seems equally plausible.

Myf-5 is part of a regulatory network in human muscle development

The development of specialized cells in higher organisms is most likely the result of the ordered activation (or inactivation) of a complex set of regulatory genes. For muscle development, two genes—the human *myd* (Pinney *et al.*, 1988) and the mouse MyoD1 (Davis *et al.*, 1987)—have been shown to be implicated in a possible hierarchy of myogenic control genes. The sequential expression of *myd* preceding the expression of MyoD1 in transfection experiments (Pinney *et al.*, 1988) led to the model of a dependent pathway for myogenic conversion of 10T1/2 cells. According to this model, the activation of *myd*, for example by DNA hypomethylation, would result in the expression of MyoD1 which in turn prepares myoblasts to differentiate in response to depletion of mitogens from the growth medium. Since the human *myd* gene has not been isolated yet and therefore the gene product is also unknown, it is interesting to speculate whether the Myf-5 cDNA could be the transcript of the *myd* gene. The following experimental criteria are available: Myf-5 appears to be functionally equivalent to *myd* and MyoD1 in its capacity to convert 10T1/2 cells to myogenic lineages; Myf-5 and mouse MyoD1 are structurally related, whereas human *myd* DNA was undetectable by DNA hybridization with the mouse MyoD1 probe (Pinney *et al.*, 1988); Myf-5 cDNA hybridizes to fragments in human genomic DNA other than those detected by cross-hybridization with the mouse MyoD1

probe. Similar information for *myd* is not available. We would like to suggest that Myf-5 represents a member of a family of human myogenic determination or *trans*-activating factor genes which are structurally related to mouse MyoD1. Myf-5 nevertheless is distinct from MyoD1 and is possibly also different from the gene product of the putative *myd* gene. This view is supported by the fact that Myf-5 seems to act rather late in the cascade of regulatory events since it is capable of *trans*-activating muscle-specific promoters in a short-term transient expression assay in non-muscle 10T1/2 cells either directly or by virtue of prior activating muscle-specific transcription factors. On the other hand, Myf-5, like *myd*, activates the endogenous MyoD1 gene in the mouse 10T1/2 cells and therefore might act upstream of MyoD1.

The observation that Myf-5-specific mRNA is present in terminally differentiated muscle tissues at very low concentrations compared to MyoD1 levels in the committed 10T1/2 myoblasts (Davis *et al.*, 1987; and our results) probably suggests that Myf-5 by itself is not a muscle-specific transcription factor but rather is involved in the myogenic commitment process and the maintenance of the muscle phenotype. It is also conceivable that Myf-5 exerts a pleiotropic effect and combines several activities in the same molecule. Interestingly, Myf-5 is not the only MyoD1-like mRNA present in human muscle tissues: at least two other transcripts, Myf-3 and Myf-4, of moderate similarity to mouse MyoD1 exist. Each of these three different transcripts is the product of a distinct human gene. The functional relationship between these genes and their involvement in myogenic determination and differentiation remains to be investigated.

Materials and methods

Isolation and sequencing of human Myf cDNA clones

A representative cDNA library from human fetal skeletal muscle was constructed using the expression vector 'λ gt11' as described previously (Seidel *et al.*, 1988). Approximately 3×10^6 plaques were screened on duplicate filters with ^{32}P-nick-translated mouse MyoD1 probe (sp. act. 1×10^8 c.p.m./μg) under standard conditions (Maniatis *et al.*, 1982). The final washes were performed in 2 × SSC (0.3 M sodium chloride, 0.03 M sodium citrate) 0.1% SDS at 50°C for 4 × 10 min. Isolated recombinant phages were purified to homogeneity, DNA was isolated and digested with *Eco*RI and the inserts were analysed according to Southern (1975) as shown in Figure 1. Inserts of selected clones Myf-3, Myf-4 and Myf-5 were subcloned into pBS plasmids (Stratagene) or M13 mp18 and mp19 phage vectors for further analysis. Partial restriction mapping was done with enzymes of BRL, New England Biolabs and Boehringer as recommended by these suppliers. The nucleotide sequence of clone Myf-5 was determined according to Sanger *et al.* (1977) on subfragments indicated in Figure 1.

Southern blot hybridization

Human and mouse genomic DNA was digested with several restriction enzymes (Boehringer), run on 0.7% agarose gels and transferred to nitrocellulose according to Southern (1975). The mouse MyoD1 probe was purified from plasmid pVZC 11 alpha by digestion with *Eco*RI. The Myf-5-specific probe represents a 250 nt-long, 3'-terminal *Pvu*II–*Eco*RI fragment, the 80 nt Myf-4-specific probe was generated by digestion with *NCI1*/*Eco*RI; the Myf-3-specific probe contains the entire 0.8 kb *Eco*RI insert of the gt11 clone Myf-3. Hybridizations were performed in 6 × SSC, 0.5% SDS, 5 × Denhardt's solution 0.01 M EDTA and 100 μg/ml of denatured salmon sperm DNA. Filters were washed under high stringency in 0.1 × SSC, 0.1% SDS at 65°C for at least 30 min.

Expression of Myf-5 cDNA

The 1.5 kb cDNA insert of Myf-5 was isolated from the corresponding gt11 clone by digestion with *Eco*RI and purified by gel-electrophoresis on 1% agarose. The fragment was cloned into *Eco*RI-digested pEMSV-scribe vector which was kindly supplied by A.Lassar. This expression vector is a derivative of EMSV-33 (Harland and Weintraub, 1985) containing the Moloney sarcoma virus LTR and SV40 poly(A) addition signal. This construct allows unique *Eco*RI insertion between the LTR and poly(A) signal sequence. The correct orientation of the Myf-5 insert in clone pEMSV-Myf 5 and the opposite orientation in clone pEMSV-con was determined by appropriate restriction mapping.

Cell cultures and DNA transfections

C3H 10T1/2, clone 8/9 and the MC-transformed (methycholanthrene) derivative MCA, clone 15 were originally isolated by C.Heidelberger (Reznikoff *et al.*, 1973) and provided to us by H.Marquardt. All cells were grown in BME medium (Gibco) supplemented with 10% fetal calf serum. Myogenic differentiation was induced by incubation in Dulbecco's modified Eagle's medium (DMEM) with 2% horse serum. To obtain G418-resistant colonies 0.5–1 μg of supercoiled pSV2-neo plasmid was co-transfected with 15–30 μg of the expression vehicle pEMSV-Myf 5 or the control plasmid pEMSV-con. For each experiment 5×10^5 cells were plated on 10 cm diameter dishes (or T-75 flasks) 24 h prior to transfection and refed 3 h before the addition of DNA. Calcium phosphate precipitate (Graham and von der Eb, 1973) was formed with premixed DNA as described previously (Lohse and Arnold, 1988). The DNA was removed 24 h later by the addition of fresh medium, supplemented with 0.4 μg/ml G418 (Geneticin, Gibco). After the appearance of macroscopically visible colonies, the medium was switched to DMEM containing 2% horse serum and 0.2 mg/ml G418. To score for myogenic colonies, plates were fixed 5 days after shifting to differentiation-supporting medium. Plates were washed three times with phosphate-buffered saline (PBS), incubated for 10 min in ice-cold methanol, again washed with PBS, stained with Giemsa and air-dried. For immunostaining cells were incubated with anti-skeletal myosin rabbit serum (Bio-Yeda, Rehovot, Israel) in a 1:20 dilution for 30 min at 32°C. After extensive washing with PBS, the second antibodies, anti-rabbit-IGG coupled to horseradish peroxidase or fluoresceital, was added, incubated for 30 min at 37°C and washed free of unbound antibody with PBS. The peroxidase reaction was performed with chlornaphthole and positive cells were scored at 20–50× magnification in the bright field microscope. Controls were treated with second antibody alone.

Transient transfection of pEMSV-Myf 5 plasmid with reporter CAT constructs in 10T1/2 fibroblasts

Approximately 5×10^5 fibroblasts were transfected on 25 cm^2 flasks with double CsCl purified, supercoiled plasmid DNA. CAT activity was determined as described by Gorman (1985). For the co-transfection experiments, 10 μg of pEMSV-Myf5 and 10 μg of one of the following plasmids were mixed and precipitated with calcium phosphate as described elsewhere (Arnold *et al.*, 1988): LC-39 and LC-64 (Arnold *et al.*, 1988); cardiac actin CAT (Quitschke *et al.*, 1987), β-actin Pl-CAT (Lohse and Arnold, 1988). Cell extracts were prepared 4 days after the transfection and CAT activities were determined in aliquots either standardized to constant percentage of cellular extract or constant protein concentration in the assay. Both calibration procedures gave the same relative values.

Acknowledgements

We thank Andrew Lassar for providing the plasmids pVZ-MyoD1 and pEMSV scribe which made the isolation of Myf-5 possible; H.Marquardt for the C3H 10T1/2 and MCA CL15 cells; H.Eberhardt for skilful technical assistance; F.Seitz for the computer analysis; C.Mink and K.D.Soehren for typing this manuscript and help with the artwork. We also gratefully acknowledge the critical comments on the manuscript by M.Buckingham. We acknowledge the postdoctoral fellowship of the Deutsche Muskelschwundhilfe e.V. supporting T.B. This work was supported by a grant to H.A. from the Deutsche Forschungsgemeinschaft.

References

Alonso,M.C. and Cabrera,C.V. (1988) *EMBO J.*, **7**, 2585–2591.

Arnold,H.H., Tannich,E. and Paterson,B.M. (1988) *Nucleic Acids Res.*, **16**, 2411–2429.

Berg,J.M. (1986) *Science*, **232**, 485–487.

Braun,T., Tannich,E., Buschhausen-Denker,G. and Arnold,H.H. (1988a) In Kedes,L. and Stockdale,T. (eds), *Cellular and Molecular Biology of Muscle Development*. UCLA Symposia on Molecular and Cellular Biology, New Series, Vol. 87. Alan R. Liss, New York, in press.

Braun,T., Buschhausen-Denker,G., Tannich,E. and Arnold,H.H. (1988b) *FEBS Lett.*, **239**, 309–312.

Cabrera,C.V., Martinez-Arias,A. and Bate,M. (1987) *Cell*, **50**, 425–433.

Cavener,D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.
Davis,R.L., Weintraub,H. and Lassar,A.B. (1987) *Cell*, **51**, 987–1000.
Gorman,C.M. (1985) In Glover,D.M. (ed.), *DNA Cloning II*. IRL Press, Oxford, UK, pp. 143–190.
Graham,F.L. and van der Eb,A.J. (1973) *Virology*, **52**, 456–467.
Harland,R. and Weintraub,H. (1985) *J. Cell Biol.*, **101**, 1094–1099.
Jones,P.A. and Taylor,S.M. (1980) *Cell*, **20**, 85–93.
Kaddurah-Daouth,R., Greene,J.M., Baldwin,A.:Jr and Kingston,R. (1987) *Genes Dev.*, **1**, 347–357.
Kingston,R.E., Baldwin,A.S. and Sharp,P.A. (1984) *Nature*, **312**, 280–282.
Konieczny,S.F. and Emerson,C.P.J. (1984) *Cell*, **38**, 791–800.
Konieczny,S.F., Baldwin,A.S. and Emerson,C.P.J. (1985) In Emerson,E., Fischman,D.A., Nadal-Ginard,B. and Siddiqui,M.A.Q. (eds), *Molecular Biology of Muscle Development*. UCLA Symposium on Molecular and Cellular Biology, New Series, Vol. 29. Alan R. Liss, New York, pp. 21–34.
Lassar,A.B., Paterson,B.M. and Weintraub,H. (1986) *Cell*, **47**, 649–656.
Lech,K., Anderson,K. and Brent,R. (1988) *Cell*, **52**, 179–184.
Lohse,P. and Arnold,H.H. (1988) *Nucleic Acids Res.*, **16**, 2787–2803.
Maniatis,T., Fritsch,E.F. and Sambrook,J.C. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Pinney,D.F., Pearson-White,S.M., Konieczny,S.F., Latham,K.E. and Emerson,C.P.,Jr (1988) *Cell*, **53**, 781–793.
Quitschike,W., Eldridge,J.D. and Paterson,B.M. (1987) In Cullen,B., Gage,L.P., Siddiqui,M.A.Q., Skalka,A.M. and Weissbach,H. (eds), *Mechanisms of Control of Gene Expression*. UCLA Symposium on Molecular and Cellular Biology, New Series, Vol. 67. Alan R. Liss, New York.
Reznikoff,C.A., Brankow,D.W. and Heidelberger,C. (1973) *Cancer Res.*, **33**, 3231–3238.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Seidel,U., Bober,E., Winter,B., Lenz,S., Lohse,P., Goedde,H.W., Grzeschik,K.H. and Arnold,H.H. (1988) *Gene*, **66**, 135–146.
Shaw,G. and Kamen,R. (1986) *Cell*, **46**, 659–667.
Shuin,T., Billings,P.C., Lillehaug,J.R., Patierno,S.R., Roy-Burman,P. and Landolph,J.R. (1986) *Cancer Res.*, **46**, 5302–5311.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Stone,J., De Lange,T., Ramsay,G., Jakobovits,E., Bishop,J.M., Varmus,H. and Lee,W. (1987) *Mol. Cell. Biol.*, **7**, 1697–1709.
Taylor,S.M. and Jones,P.A. (1979) *Cell*, **17**, 771–779.
Thisse,B., Stoetzel,C., Gorostiza-Thisse,C. and Perrin-Schmitt,F. (1988) *EMBO J.*, **7**, 2175–2185.
Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5205.
Villares,R. and Cabrera,C.V. (1987) *Cell*, **50**, 415–424.

Received on December 2, 1988; revised on December 19, 1988