Nucleic Acids Research, Vol. 19, No. 20 5645–5651

The four human muscle regulatory helix-loop-helix proteins Myf3 – Myf6 exhibit similar hetero-dimerization and DNA binding properties

Thomas Braun and Hans Henning Arnold*
Department of Toxicology, Medical School, University of Hamburg, Grindelallee 117, 2000 Hamburg 13, FRG

Received July 8, 1991; Revised and Accepted September 20, 1991

ABSTRACT

The muscle regulatory proteins Myf3, Myf4, Myf5, and Myf6 share a highly conserved DNA binding and dimerization domain consisting of a cluster of basic amino acids and a potential helix-loop-helix structure. Here we demonstrate that the four human muscle-specific HLH proteins have similar DNA binding and dimerization properties. The members of this family form protein complexes of comparable stability with the ubiquitously expressed HLH proteins E12, E2-2, and E2-5 and bind to the conserved DNA sequence CANNTG designated as E-box with similar efficiency *in vitro*. The binding affinities of the various complexes are greatly influenced by the variable internal and flanking nucleotides of the consensus motif. Combinations of Myf proteins with one another and with lyl-1, an HLH protein from human T cells, do not bind to DNA *in vitro*. Our results suggest that combinatorial associations of the various tissue-specific and more widely expressed HLH factors do not result in differential recognition of DNA sequences by Myf proteins.

INTRODUCTION

Myogenesis is a complex biological process which involves determination of multipotential mesodermal stem cells to the myogenic lineage and their subsequent differentiation into functional myocytes. During the transition of proliferating muscle progenitor cells to terminally differentiated muscle cells *in vivo*, distinct fiber types are established, probably in response to developmental and physiological signals. Each of the developmental steps is characterized by the transcriptional activation of specific sets of genes, most notably during terminal differentiation when the muscle isoforms of contractile proteins are expressed. The coordinate activation of cell type-specific and developmentally regulated genes suggests that a myogenic program exists which controls changes in gene activities. Although the molecular mechanisms by which these regulatory events take place remain largely unknown, they probably involve transactivating proteins which recognize control regions in muscle genes. Indeed, regulatory DNA elements that confer muscle-specific expression and serve as binding sites for putative muscle transcription factors have been identified in multiple muscle genes (for review see ref. 1 and 2).

Important clues towards a better understanding of transcriptional control during myogenesis came from the discovery of a family of muscle-specific nuclear proteins which are capable to convert C3H mouse 10T1/2 fibroblasts to the muscle phenotype (for review see ref. 2). MyoD1, the prototype of these myogenic factors, was first cloned from a mouse muscle cell line (3). Subsequently the related but distinct proteins myogenin (4,5), Myf5 (6), and MRF4/herculin/Myf6 (7,8,10) have been identified. The cDNAs encoding each of the four factors from humans have been described (6,9,10). These muscle regulatory proteins share a conserved sequence of a cluster of basic amino acids preceding a putative helix-loop-helix (HLH) domain. This common motif is also present in other regulatory proteins including those controlling developmental processes in drosophila (11 – 14) and in the human immunoglobulin enhancer-binding proteins E12, E47 (15,16), and E2-2 (17).

The conserved HLH domain allows members of this family to form oligomeric protein complexes (15,16,17,18) which bind to DNA with high affinity via their basic amino acid regions (19,20). The DNA binding site is represented by the consensus sequence CANNTG, known as E-box, which is part of many muscle and non-muscle genes. In particular, this sequence is present in the tissue-specific enhancers of the muscle phosphocreatine kinase (MCK) gene (21,22), the myosin light chain 1/3 gene (23,24), and the troponin I gene (25) as well as in the promoter of the α-actin (26) and α-acetylcholine receptor genes (27). Muscle-specific HLH proteins can bind to these control regions *in vitro* (19,20,28,10), and probably play an important role in their tissue-specific expression. In fact, it has been demonstrated that reporter genes under the control of muscle-specific enhancers can be transactivated in fibroblasts by the overexpression of the myogenic regulators (5,6,9,10, 29,30,31). Dimerization of tissue-specific and ubiquitously expressed HLH proteins to form DNA binding complexes, leads

* To whom correspondence should be addressed

to the intriguing idea that differential gene control may involve combinatorial associations of the various HLH proteins. Despite the general knowledge on fundamental structural features and some functional aspects of the muscle regulatory proteins and their ubiquitously expressed counterparts, detailed information on their biochemical properties are lacking.

In an attempt to define some of the distinct biochemical features we compared the capacity of the four human Myf proteins to form hetero-oligomers with the ubiquitous human HLH proteins E12, E47, E2-2 and with lyl-1 and studied the specificity with which the different complexes bind to DNA. Here, we report hat DNA binding occurs only in defined combinations of HLH proteins and that all hetero-oligomeric Myf complexes exhibit similar DNA binding properties. The ability of the Myf proteins to interact with the ubiquitous E-type HLH proteins but not with each other may be part of a mechanism that selectively controls gene expression in different tissues. There is however no evidence for selectivity in protein-protein interactions or in DNA binding among the four human muscle regulators Myf3,4,5, and 6.

MATERIALS AND METHODS

Construction of plasmids

To synthesize RNA templates for *in vitro* translation, the following plasmids were constructed. The human Myf5 cDNA was digested with AvaI and EcoRI, blunt ended with T4 DNA polymerase and fused in frame to the initiation codon of the vector pT7βSal (32). The complete coding sequence for Myf4 was obtained by fusing the NcoI-SstI genomic fragment representing the 5'end of exon 1 to the SstI-EcoRI fragment of the Myf4 cDNA. The combined construct was ligated into the NcoI site of the plasmid pT7βSal. The full length cDNA for Myf6 (10) was cloned as EcoRI fragment into the plasmid pBS. The full length cDNA for Myf3 (kindly provided by C.Emerson and S.Pearson-White) was cloned as EcoRI fragment into the pBS vector. Plasmid pE12R, containing a cDNA fragment of E12 in the pBS-ATG vector was obtained from C. Murre (16). The plasmids pT7E2-2 and pT7E2-5 containing the full length cDNAs for E2-2 and E2-5 in the vector pT7βSal were obtained by P.Henthorn (17).

In vitro transcription and translation

mRNA templates were synthesized from the respective plasmid constructs digested with restriction enzymes to linearize the vectors downstream of the coding region. Transcription of RNA was performed in 100 µl reactions using T3 or T7 RNA-polymerase. The newly synthesized RNA was purified by extractions with phenol and chloroform and precipitated with ethanol prior to use in the *in vitro* translation reaction. mRNA templates were translated in 50 ul of pretreated rabbit reticulocyte lysate (Promega) in the presence of S^{35}-methionine according to the manufacturer's recommendations. Aliquots were analyzed on SDS-polyacrylamide gels and subjected to autoradiography to estimate the amount of synthesized proteins. Concentrations of proteins were estimated from the radioactivity incorporated per number of methionine residues per protein.

Synthetic DNA binding sites

The oligonucleotides MLCA (LC-enh) GATCAAGTAACAGC-AGGTGCAAAATAAAGT and MLCB GATCCATCTACACC-AGCTGGCAAAAATGAC were synthesized according to the sequences present in the human MLC1/3 enhancer (23). The oligonucleotide LC-Mut is a derivative of MLCA and contains the following sequence: GATCAAGTAAGTAACTGTGC AAAATAAAGT (the mutated nucleotides are underlined). Oligonucleotides EMB-1 and EMB-2 were derived from the promoter region of the human MLC1emb gene and have the sequences GATCAATACACAGTTGTCAGCTGTACCTGCT and GATCCTTTTATAGTCAGCAGCAGTTGCTCTT, respectively (31). The purification of synthetic oligonucleotides and their radiolabelling for band shift experiments have been described previously (34).

Electrophoretic mobility-shift assay (EMSA)

The binding reactions were performed in a buffer containg 10 mM Tris pH 7.5, 50 mM NaCl, 1 mM DTT, 1 mM EDTA and 5% glycerol with aliquots of reticulocyte lysate containing the translated proteins. Usually 2 to 12 µl of lysate were used in a total reaction volume of 25 µl containing 50000 cpm of labelled fragment and 3 µg poly dI/dC as unspecific competitor. Following incubation for 30 min at room temperature, DNA protein complexes were resolved on 3.5% polyacrylamide gels in 0.25×TBE running buffer. For more stringent conditions, band-shifts were performed in the presence of the non-ionic detergent NP40 (final concentration 0.5%) and 300 ng sheared salmon sperm DNA and gels were run in 1×TBE buffer. To shift DNA-protein complexes with antibodies, binding reactions were preincubated for 20 min at room temperature 1−3 µl antibody dissolved in PBS was added and incubations were continued for additional 20 min.

Antibodies

Polyclonal serum to Myf5 and the monoclonal antibody hmyf4-21.6 was raised against the human Myf5-glutathione transferase fusion protein (10). A full description of the monoclonal Myf5 antibody generated in collaboration with M.Buckingham and the Hybrido Laboratory at Institut Pasteur, Paris, is in preparation. The subclone anti-hmyf5-21.6 used in the present study is a IgG2b immunoglobulin and was used for ascites production in mice to yield high titer antibodies. The IgG fraction was precipitated from ascites fluid with ammonium sulfate and dialyzed against several changes of PBS before use in band-shift experiments.

Antiserum against E12 was kindly supplied by C.Murre and D.Baltimore. The mouse monoclonal antibody against rat myogenin was a kind gift of Woody R.Wright, Dallas, Texas. The monoclonal antibodies against MyoD were generously supplied by S.Kohtz, New York.

Methylation interference footprinting

Single stranded synthetic oligonucleotides encompassing the E-box binding site were radioactively labelled with polynucleotide kinase. Following inactivation of the kinase, the complementary strands were annealed and partially methylated with DMS for 1 min on ice in the presence of 2 µg carrier DNA. Methylated double stranded DNA was recovered by ethanol precipitation. Labelled fragments were purified on a non-denaturing 5% polyacrylamide gel before use in binding reactions. For DNA/protein complex formation an analytical EMSA reaction was scaled up 5 times. The entire reactions were run on 3.5% non-denaturing polyacrylamide gel. The gel was exposed to x-ray film for 2 h at 4°C, bound and free DNA probes were excised and the DNA was electro-eluted. The eluted DNA was extracted twice with phenol and chloroform and precipitated with ethanol.

The pellet was dissolved in 100 µl of a 1% piperidine solution, incubated for 30 min at 90°C and lyophilized twice before the cleavage products were analyzed on a 10% sequencing gel.

Immunoprecipitation

Proteins were translated *in vitro* in the presence of ^{35}S-methionine and preincubated for 15 min at 37°C for immunoprecipitation. Then 1 volume (25 µl) of solubilization buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% NP40) and 2−4 µl antiserum were added and the reaction was incubated for 1 h at 4°C on a rotating platform. 35 µl of proteinA-sepharose beads equilibrated in buffer A (10 mM Tris pH 7.5, 150 mM NaCl, 0.2% NP40, 2 mM EDTA) were added and incubation was continued for another hour. The beads were washed 3 times in buffer A and 2 times in buffer B (10 mM Tris pH 7.5, 500 mM NaCl, 0.2% NP40, 2 mM EDTA), collected by centrifugation and bound protein as recovered by boiling in SDS-sample buffer. The proteins were separated on a 12% SDS-polyacrylamide gel. For fluorography the gel was soaked in 1 M sodium salicylate for 1 h before drying.

RESULTS

Myf factors associate with E12 but not with other tissue-specific HLH proteins to bind to DNA with high affinity

The binding properties of the four human myogenic factors have been analyzed in gel mobility shift assays on the synthetic oligonucleotide MLCA. This sequence was derived from the tissue-specific enhancer of the human MLC1/3 gene which is highly conserved to the analogous element of the rat MLC1/3 gene (23,24) and the enhancer of the MCK gene (21,22) and confers muscle-specific expression to CAT reporter constructs. The proteins were produced by cell free translation of the appropriate mRNA templates in rabbit reticulocyte lysate and were incubated with the radiolabelled DNA binding site. As shown in Figure 1A, using constant amounts of E12 and increasing concentrations of Myf proteins ranging from one third subequimolar to equimolar concentrations in relation to E12, the homomeric E12 DNA complex which bound to DNA under these conditions (fig. 1A, lanes 1,5,9 and 13) was readily displaced by the E12/Myf hetero-oligomers. This was evident from the gradual disappearance of the larger E12 complex and the reciprocal accumulation of the faster migrating aggregates containing the smaller Myf proteins. Whether the preferred formation of E12/Myf DNA complexes is due to higher DNA affinity of the hetero-oligomers or to the preference of Myf/E12 hetero-oligomerization over E12 homo-dimerization can not be decided from this experiment.

To confirm the presence of both, E12 and Myf HLH proteins as part of the DNA binding complex, specific antibodies directed against the individual Myf proteins and E12 were applied. As shown in Figure 1B, the addition of polyclonal antiserum against human Myf5 and monoclonal antibody against rat myogenin (Myf4) resulted in distinct supershifts which were represented by the larger complexes containing the IgG proteins. The antibody to mouse MyoD (Myf3) as well as the antiserum to E12 did not generate supershifts but instead interrupted DNA binding of the complexes. Antibody to Myf6 was not available.

To test the ability of other HLH factors to participate in sequence-specific DNA binding, lyl-1 protein originally identified in human leucemia cells (35), was mixed with each of the Myf proteins or E12 and subjected to the gel mobility-shift assay. In

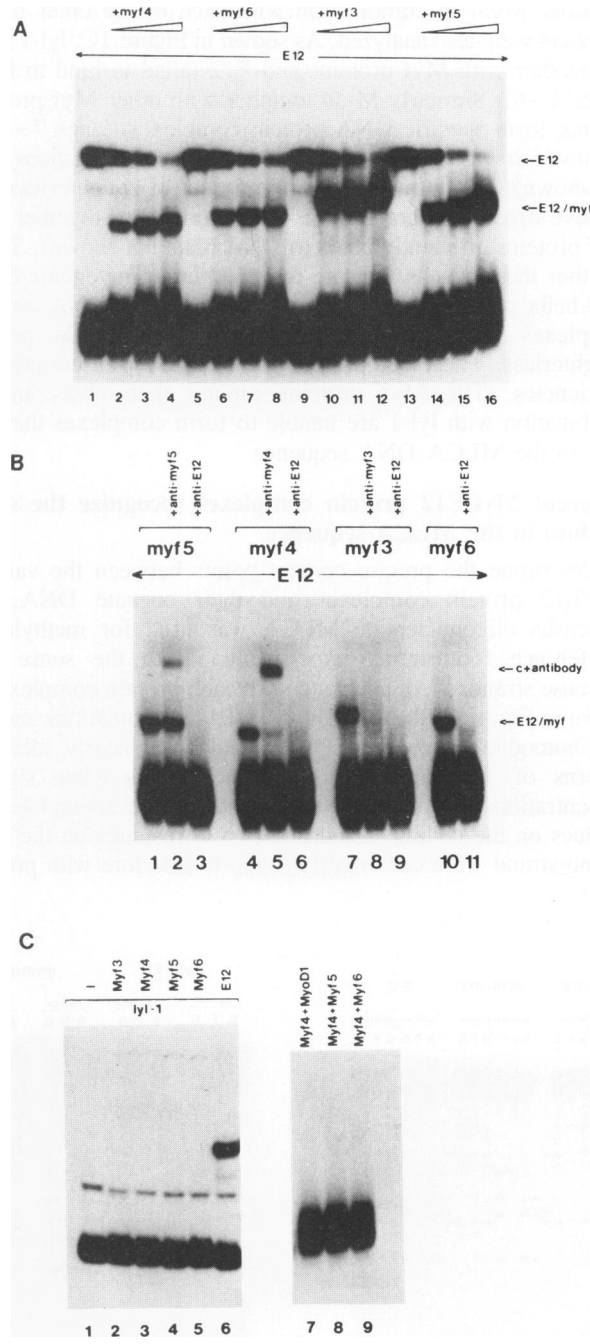

Figure 1: DNA binding complexes of the human Myf proteins with E12. A) Electrophoretic mobility-shift assays (EMSA) with increasing concentrations of Myf proteins (1/3, 2/3 and 3/3 of E12 concentrations) and constant amounts of E12 were performed with the radiolabelled oligonucleotide MLCA (AGTAA-CAGCAGGTGCAAAATAAAGT). Concentrations of the *in vitro* translated proteins were estimated from the incorporation of sulphur-35 per number of methionine residues present in each protein. Details for the *in vitro* translation in reticulocyte lysate, and the conditions for binding and electrophoresis are given in Experimental Procedures. Specific complexes are indicated, the intense, fast moving complexes are non-specific associations derived from the reticulocyte lysate. B) Antibody supershifts or suppression of Myf/DNA binding complexes. EMSAs were performed with polyclonal antibodies to Myf5 (lane 2) and E12 (lanes 3,6,9 and 11), and monoclonal antibodies to myogenin (Myf4) (lane 5) and MyoD1 (Myf3) (lane 8). C) EMSAs with *in vitro* synthesized, Myf proteins, E12, and lyl-1 (lanes 1−6), and with mixtures of Myf4 and individual members of the Myf family (lanes 7−9). The E12 complex in lane 6 represents the homo-dimer. The fast migrating complexes are due to unspecific binding of components in the reticulocyte lysate.

addition, Myf4 in combination with each of the other family members were also analyzed. As shown in Figure 1C, lyl-1 alone or together with Myf proteins and E12 failed to bind to DNA (lanes 1–6). Similarly Myf4 together with other Myf proteins did not form specific DNA protein complexes (lanes 7–9), a result which was also obtained with other Myf combinations (data not shown). In contrast, the gene product of daughterless, the putative drosophila homologue of E12, alone or together with Myf proteins efficiently binds to DNA (data not shown). Taken together these results indicate that the human myogenic helix-loop-helix proteins Myf3 to Myf6 readily form DNA binding complexes with E12 and the related drosophila protein daughterless. These complexes bind to DNA with comparable efficiencies. The Myf proteins among themselves and in combination with lyl-1 are unable to form complexes that can bind to the MLCA DNA sequence.

Different Myf/E12 protein complexes recognize the same residues in the MLCA sequence

To determine the precise contact points between the various Myf/E12 protein complexes and their cognate DNA, the consensus oligonucleotide MLCA was used for methylation interference footprinting experiments. Both the sense and antisense strands were analyzed with each protein complex. As shown in Figure 2, the different Myf/E12 combinations and the E12 homodimer bound to DNA resulted in nearly identical patterns of suppressed G (guanosine) bands when similar concentrations of proteins were used. Three methylated G residues on the coding strand and two G residues on the non-coding strand were consistently found to interfere with protein binding. Marginal, if any differences appeared in the residual intensities of the reduced bands between the various protein complexes. This indicates that, under the applied *in vitro* binding conditions, the four human muscle regulatory factors Myf3, Myf4, Myf5, and Myf6 and the ubiquitous HLH protein E12 recognize the same nucleotides in the MLCA consensus sequence.

To ascertain the specificity of the protein interactions with DNA, the oligonucleotide LC-Mut containing several mutated nucleotide residues in the conserved recognition sequence was subjected to gel mobility-shift assays in parallel to the wild type oligonucleotide LC-enh (identical to MLCA). As seen in figure 3, Myf3, Myf4, Myf5, and Myf6 each cotranslated with E12 protein formed the specific DNA binding complexes C2 on the wild type oligonucleotide while complex formation was generally reduced on the mutant oligonucleotide. The complex C1 was caused by components of the reticulocyte lysate. This result was corroborated in competition experiments with unlabelled wild type or mutant oligonucleotides used in 20 fold molar excess over the radio-labelled binding site. The wild type competitor WT prevented complex formation, the mutant sequence MT did not. There was no apparent difference in the specificity of DNA recognition between the four Myf proteins.

The non-conserved residues in the MyoD binding site influence binding affinity

Based on the compilation of E-box sequences within cis-acting elements conferring tissue-specific expression to muscle genes, the general consensus sequence NCANNTGN has been proposed for the MyoD1 binding site (36). In the human MLC1/3 core enhancer and the promoter of the human MLC1emb gene several sequence elements follow this rather degenerate consensus motif (23,33). We have utilized these naturally occurring variants of the putative binding site to compare the efficiency of complex

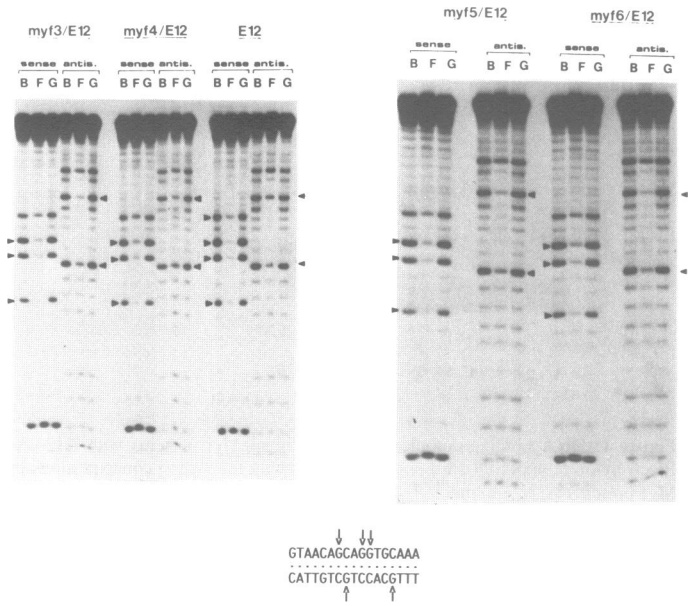

Figure 2: Methylation interference footprints on the MLCA binding site generated by *in vitro* translated Myf and E12 complexes. The 5'end-labelled oligonucleotide MLCA was incubated with the indicated protein complexes and subjected to footprinting as described in Experimental Procedures. Arrowheads mark the methylated residues which interfere with protein binding as indicated by the reduced intensity in the unbound DNA (F) versus bound DNA (B). The guanosine sequencing ladder is shown as G lane. Only the part of the MLCA oligonucleotide sequence is shown that illustrates the contact points for the protein complexes on both strands (arrows).

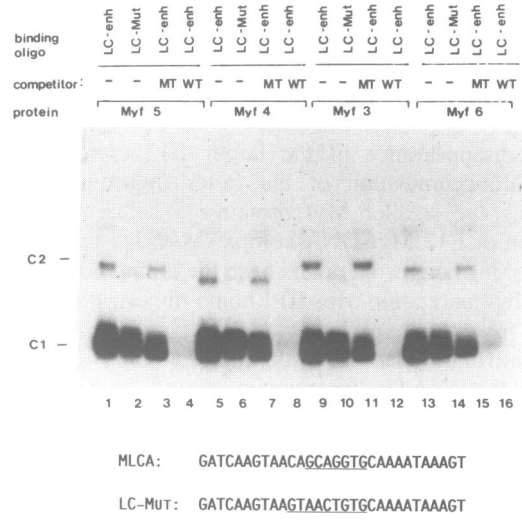

Figure 3: The human Myf/E12 complexes bind to the conserved E-box DNA sequence but not to a mutated version of the consensus. EMSAs were performed with the indicated proteins synthesized in reticulocyte lysate (50 μl) and the radiolabelled double-stranded oligonucleotides LC-enh (sequence is shown as MLCA) and LC-Mut. Competitions were carried out in the presence of 20 fold molar excess of unlabelled oligonucleotide LC-enh (WT=MLCA) or LC-Mut (MT). Complex C1 is due to components in the reticulocyte lysate (data not shown). Complexes C2 indicate the specific Myf/DNA complexes.

formation with Myf proteins. As shown in Fig. 4, gel mobility-shifts with oligonucleotides MLCA and MLCB, and EMB-1 and EMB-2 encompassing the respective consensus motifs in the human MLC1/3 enhancer (23) and the human MLC1emb promoter (31,33), resulted in strong *in vitro* DNA binding of all four Myf proteins on oligonucleotide MLCA, but considerably weaker interactions on the other three sequences. Exemplified on four different versions of the proposed DNA binding site, our results suggest that the nature of the internal nucleotides and the immediately adjacent external nucleotides of the core consensus influence the binding affinity for the muscle-specific HLH proteins.

The HLH proteins E2-2, E2-5 and E12 support high affinity DNA binding of Myf proteins but do not modulate the binding specificity

Two gene products of the human E2A gene, namely E12 and E47 (15,16), the latter was also called E2-5 (17), have been described. In addition, the widely expressed HLH protein E2-2 was recently identified (17). To explore their ability to participate in complex formation and DNA binding of the human Myf factors and to analyze their possible modulating effect on the binding specificity, we performed gel mobility-shift assays with all possible combinations of Myf proteins and the ubiquitously expressed HLH proteins of the E-type. To increase the stringency of the binding conditions, the non-ionic detergent Nonidet P40 (0.5%) and higher salt concentrations in the gel running buffer were applied. As shown in Fig. 5A, each of the individual Myf proteins bound to the oligonucleotide MLCA in hetero-oligomeric complexes formed with each one of the three E-type HLH proteins. Under the applied conditions homodimers of E2-2 and E12 did not bind, in contrast to E2-5 which appeared to bind as efficiently as the hetero-oligomers (lanes 1, 6 and 11). The reduced intensity of the unspecific band in lanes 1, 6 and 11 are due to the lesser amount of reticulocyte lysate used in these assays. A similar pattern was also observed with the drosophila protein daughterless in the absence and presence of the Myf proteins (data not shown).

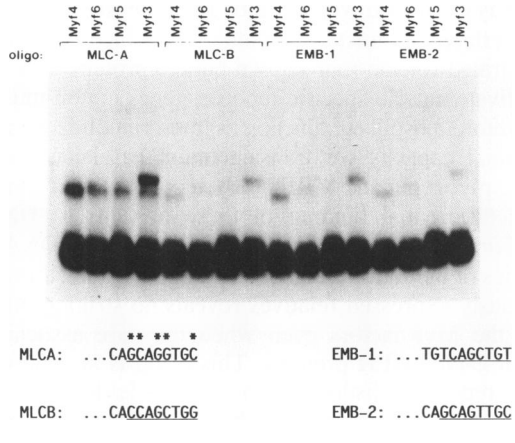

Figure 4: The internal and flanking nucleotides in the conserved sequence *C-ANNTG* influence DNA binding of Myf/E12 protein complexes. ^{32}P-labelled double stranded oligonucleotides encompassing the E-box represented by MLC-A, MLCB, EMB-1, and EMB-2 were used to perform EMSA with *in vitro* synthesized Myf proteins and E12. The faster migrating complex is caused by the reticulocyte lysate. The full sequences of the oligonucleotide binding sites are given in Experimental Procedures.

There was no appreciable difference in the extent to which the various protein complexes associated with DNA suggesting that neither heterodimer formation nor the affinity to the MLCA site were selectively affected by any of the ubiquitous HLH proteins. The lack of modulating influence on the specificity of DNA recognition by different E-type proteins was further corroborated in band shift experiments using the MLCB oligonucleotide as binding sequence (Fig. 5B). As already observed for the Myf/E12 complexes, (see fig.4), there was a strong reduction in DNA binding of the E2-2/Myf and E2-5/Myf complexes when the binding sequence was altered from MLCA to MLCB. However, equally weak binding was obtained on MLCB with all four Myf proteins regardless whether E2-2, E2-5 or E12 served as partner in the complex formation. These observations suggest that the recognition of DNA sequences is not different for the various Myf complexes regardless which E-type HLH protein contributes as potential half-site to the binding. In conclusion, these results and those shown in Fig. 1 indicate that specific DNA binding is favoured by certain combinations of HLH proteins. Which of the combinations effectively bind to DNA with high affinity is possibly controlled by their ability or disability to form dimeric

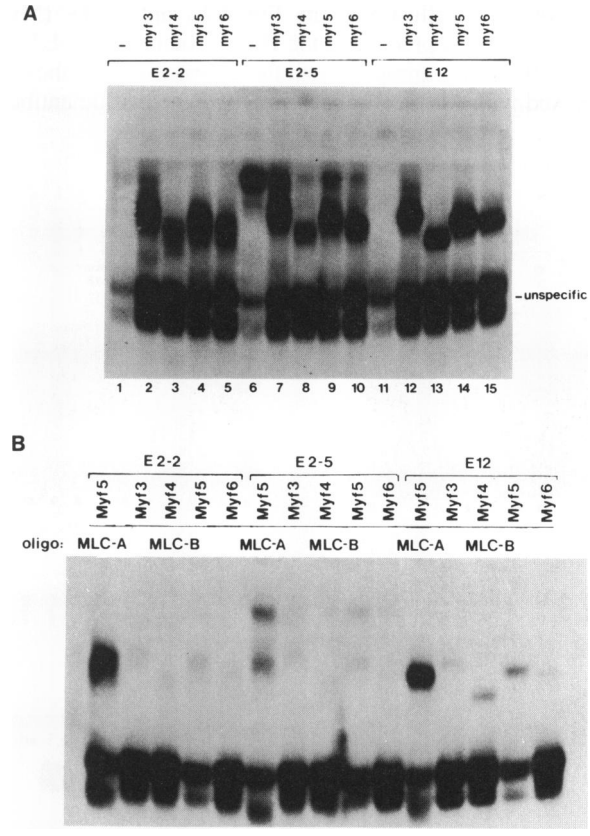

Figure 5: Complexes of Myf proteins with different ubiquitously expressed HLH proteins E2-2, E2-5, and E12 exhibit similar DNA binding properties. A) The human Myf proteins were translated *in vitro* together with E2-2 (lanes 2-5), E2-5 (lanes 7–10), and E12 (lanes 12-15) and subjected to EMSA under stringent conditions (+ 0.5% NP40) using the oligonucleotide MLCA as binding site (for complete sequence see Experimental Procedures). B) The double stranded oligonucleotide MLCB (see: Experimental Procedures) was radiolabelled and used to perform EMSA with the indicated combinations of HLH proteins translated *in vitro*. Myf5 binding to MLCA is shown for comparison. Note that binding was assayed here in the absence of NP40 to facilitate the formation of weak complexes.

or oligomeric complexes. The DNA recognition properties, however, do not seem to be different for the various protein complexes.

E12, E2-2, and E2-5 form heterodimers with the muscle-specific Myf proteins in the absence of DNA

It has been demonstrated previously that mouse MyoD1 and myogenin can form heterodimers with E12 even in the absence of DNA binding sites (16,19). To extend this investigation to all known human muscle-specific HLH proteins and to compare their relative affinites we examined their ability to form heterologous protein complexes *in vitro* with the ubiquitous HLH factors E12, E2-2 and E2-5. ^{35}S-methionine labelled *in vitro* translated products were incubated in various combinations and immunoprecipitated using antibodies which specifically recognized only one of the protein components present in the incubation mixture. The coprecipitated proteins were analyzed on SDS-polyacrylamide gels and visualized by autoradiography. Monoclonal antibodies raised against mouse MyoD1, rat myogenin, and human Myf5 as well as polyclonal serum directed against human E12 were used after they had been evaluated to be specific for the individual HLH proteins. The MyoD1 and myogenin antibodies recognized human Myf3 and Myf4 proteins, respectively. As illustrated in Fig. 6A, anti-myf3 (MyoD) antibodies failed to precipitate the HLH proteins E2-2, E2-5 and E12 but effectively brought down these proteins when they were complexed with Myf3. Similarly, myogenin specific antibodies successfully precipitated the ubiquitous E-type HLH proteins when they were associated with Myf4 protein (Fig. 6B). The relative amounts of the precipitated antigens and their associated E-type HLH proteins appeared approximately equal suggesting that both Myf3 and Myf4 readily formed stoichiometric protein complexes with each of the three ubiquitously expressed HLH proteins E2-2, E2-5 and E12. The anti-Myf5 antibody only marginally coprecipated E2-2 and E12 and slightly better E2-5 when present in heterodimers with Myf5 (Fig. 6C). It is presently unclear whether the weak coprecipitation was due to the antibody that might interfere with complex formation or alternatively whether Myf5/E12 and Myf5/E2-2 are inherently less stable. Rather weak and variable coprecipitation was also obtained with antiserum to E12 that also recognizes E2-2 and E2-5 (fig. 6 D). While Myf6, Myf5, and Myf3 when associated with E12 could be immuno-precipitated to some extent, Myf4 completely failed to be precipitated. Since Myf4-E12 complexes could be demonstrated with the Myf4 antibody, we ascribe the moderate efficiency and the large variability of immunoprecipitations with E12 antiserum to the properties of this antibody preparation rather than to different dimerization properties of the HLH proteins.

It is interesting to note that two other HLH proteins, e.g. lyl-1 and the inhibitor for DNA binding of MyoD1, Id, which are both unable to interact with the MLCA DNA binding site *in vitro*, nevertheless form heterodimers with E2-2, E2-5 and E12 (data not shown).

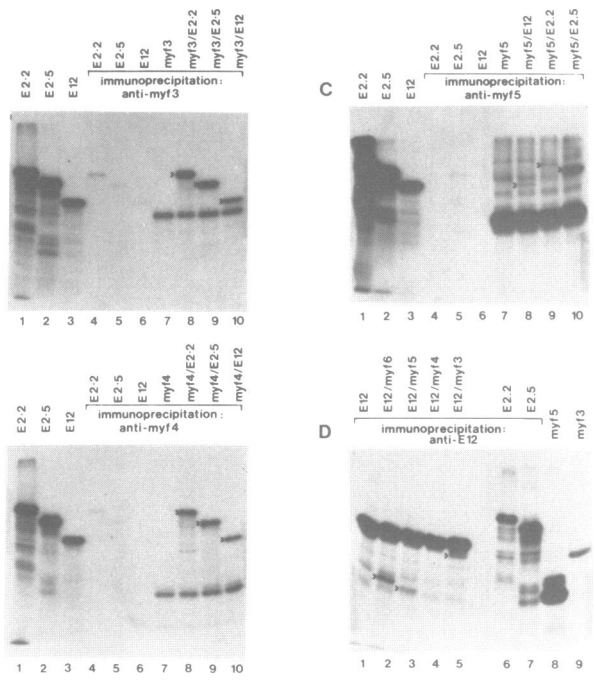

Figure 6: Myf proteins form hetero-oligomers with the widely expressed HLH proteins E2-2, E2-5, and E12 in the absence of DNA. *In vitro* synthesized and radioactively labelled E2-2, E2-5, and E12 proteins and cotranslated Myf3 (A), Myf4 (B), and Myf5 (C) were incubated for complex formation followed by immunoprecipitation with the respective anti-Myf antibodies. Precipitated proteins were analyzed on SDS polyacrylamide gels and visualized by autoradiography. Arrowheads indicate E-type HLH proteins coprecipitated by antibodies to Myf3 (A), Myf4 (B), and Myf5 (C). Antibody to E12 was utilized to coprecipitate Myf proteins complexed to E12 (D). This antibody also recognizes E2-2 and E2-5 proteins.

DISCUSSION

The discovery of a number of independent MyoD1 related muscle regulatory proteins which are capable to convert 10T1/2 fibroblasts to the muscle lineage has complicated the original concept that a single genetic locus or few linked genes would be sufficient to initiate myogenesis (37,38). In particular it has been difficult to assess the role of the individual factors in establishing the muscle phenotype in tissue culture cells since forced expression of one factor generally leads to an autoregulatory activation of the corresponding endogenous gene and the other members of the gene family (9,39). In established myogenic cell lines usually only a subset of muscle regulatory proteins is expressed which seems to argue for redundant or at least overlapping functions among these factors. It has been demonstrated for the four known human proteins that they can transactivate muscle-specific reporter genes in non-muscle cells and therefore most likely function as transcription factors (10,23). Differential capacity for transactivation has been noticed for MyoD1, myogenin and MRF4 (30), a phenomenon which could involve differential binding of these proteins to DNA. Our detailed *in vitro* analysis of heterodimerization and DNA binding properties of the four human muscle regulatory factors and their ubiquitously expressed relatives reveals no striking differences among the Myf factors even when they are associated with different E-type HLH proteins. This result is in contrast to our previous reports on distinct binding efficiencies for the glutathione fusion proteins which probably bind as homodimers and show marked differences in DNA complex formation (10,23). A possible explanation for this discrepancy could be the reduced capacity to homodimerize of some GS-Myf fusion proteins. The DNA sequence G*CA*GG*TG* appears to be the commonly recognized and most efficient binding site for the four human muscle HLH proteins regardless whether E12, E2-2, or E2-5 participates in the complex. The replacement of the internal pairs

of guanosine residues by GC or GT and/or changes of the flanking nucleotides from G to C or T strikingly diminishes binding. This observation is supported by the results of methylation interference footprinting which show that methylation of the conserved G residues interferes with the binding of E12 homodimers as well as Myf/E12 heterodimers. Our findings are also in agreement with recently obtained data on selected and amplified binding sites for MyoD, E12 and E47 (18,40).

The similar properties of the different myogenic DNA binding complexes probably reflect the structural homologies of the basic and helix-loop-helix regions of the participating proteins. Likewise, the similar heterodimerization properties may be the result of the structural conservation of both amphipathic helices. Why then do Myf proteins alone and homodimers of E2-2 or E12 so inefficiently bind to DNA? Recently it has been shown that an inhibitory domain, located N-terminal to the basic region of E12, prevents E12 homodimers but not E12/MyoD heterodimers from binding to DNA, whereas MyoD homodimerizes poorly and therefore does not efficiently form DNA complexes (18). The reason for inefficient dimerization of Myf proteins has not yet been elucidated but could be due to the fact that *in vitro* translated proteins might lack the appropriate posttranslational modifications to facilitate homodimer formation.

In conclusion, the *in vitro* experiments presented in this report provide evidence that no major differences in DNA binding and protein-protein interactions exist among the human muscle-specific factors. In particular, the recognition of the E-box sequence which serves as cis-acting element in many tissue-specific genes appears to be very similar and a modulating effect by the various E-type HLH proteins present in the binding complex could not be observed. Taken together these observations argue that all members of the Myf family may perform similar functions. On the contrary, the notion that the myogenic factors exhibit distinct expression patterns during early mouse development and in later myogenesis (41,42,43) may suggest that they serve specific biological purposes which probably involve regions of the proteins other than the conserved basic and helix-loop-helix domains.

ACKNOWLEDGEMENTS

We thank C.Murre, W.Wright and S.Kohtz for valuable reagents. We gratefully acknowledge the assistance by M.Buckingham, J.C.Mazei and F.Nato, Pasteur Institut, in preparing the monoclonal antibody to Myf5. We thank the members of the lab for helpful discussions and in particular A.Lassen and H.Eberhardt for technical assistance and A.Broecker-Nagel for her secretarial help. The work was supported by grants from Deutsche Forschungsgemeinschaft and Deutsche Muskelschwundhilfe e.V.

REFERENCES

1. Rosenthal, N. (1989) Curr.Opinion. Cell Biol. 1, 1099–1101
2. Olson, E.N. (1990) Genes Dev. 4, 1454–1461
3. Davis, R.L., Weintraub, H., and Lassar, A.B. (1987) Cell 51, 987–1000
4. Wright, W.E., Sassoon, D.A., and Lin, V.K. (1989) Cell 56, 607–617
5. Edmondson, D.G. and Olson, E.N. (1989) Genes Dev. 3, 628–640
6. Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E., and Arnold, H.H. (1989) EMBO J. 8, 701–709
7. Rhodes, S.J. and Konieczny, S.F. (1989) Genes Dev. 3, 2050–2061
8. Miner, J.H., and Wold, B. (1990) Proc.Natl.Acad.Sci USA 87, 1089–1093
9. Braun, T., Bober, E., Buschhausen-Denker, G., Kohtz, S., Grzeschik, K. and Arnold, H.H. (1989) EMBO J. 8, 3617–3625
10. Braun, T., Bober, E., Winter, B., Rosenthal, N. and Arnold, H.H. (1990) EMBO J. 9, 821–831
11. Caudy, M., Vassin, H., Branc, M., Tuma, R., Yeh-Jan, L., and Jan, Y.N. (1988) Cell 55, 1061–1067
12. Klambt, C., Knust, E., Tietze, K. and Campos-Ortega, J.A. (1989) EMBO J. 8, 203–210
13. Thisse, B., Stoetzel, C., Gorostiza-Thisse, C., and Perrin-Schmitt, F. (1988) EMBO J. 7, 2175–2183
14. Villares, R., and Cabrera, C.V. (1987) Cell 50, 415–424
15. Murre, C., McCaw, P.S. and Baltimore, D. (1989) Cell 59, 777–783
16. Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Kan, Yn.N. Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., Weintraub, H. and Baltimore, B. (1989) Cell 58, 537–544
17. Henthorn, P., Kiledjian, M. and Kadesch, T. (1990) Science 247, 467–470
18. Sun, X-H. and Baltimore, D. (1991) Cell 64, 459–470
19. Brennan, T.J. and Olson, E.N. (1990) Genes Dev. 4, 582–595
20. Davis, R.L., Cheng, P.F., Lassar, A.B., and Weintraub, H. (1990) Cell 60, 733–746
21. Sternberg, E., Spizz, G., Perry, W.M., Vizard, D., Weil, T. and Olson, E. (1988) Mol.Cell.Biol. 8, 2806–2809
22. Horlick, R.A., and Benfield, P.A. (1989) Mol.Cell Biol. 9, 2396–2413
23. Rosenthal, N., Berglund, E.B., Wentworth, B.M., Donoghue, M., Winter, B., Bober, E., Braun, T. and Arnold, H.H. (1990) Nucleic Acids Res.18, 6239–6246 - 29 -
24. Donoghue, M., Ernst, H., Wentworth, B., Nadal-Ginard, B., and Rosenthal, N. (1988) Genes Dev. 2, 1779–1790
25. Lin, H., Yutzey, K.E. and Konieczny, S.F., (1991) J.Mol.Cell.Biol. 11, 267–280
26. Sartorelli, V., Webster, K.A., and Kedes, L. (1990) Genes Dev. 4, 1811–1822
27. Piette, J., Bessereau, J.L., Huchet, M. and Changeux, J.-P., (1990) Nature 345, 353–355
28. Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D. and Weintraub, H. (1989) Cell 58, 823–831
29. Lin, A.-Y., Dechesne, C.A., Eldridge, J. and Paterson, B.M. (1989) Genes Dev. 3, 986–996
30. Yutzey, K.E., Rhodes, S.J. and Konieczny, S.F. (1990) Mol.Cell.Biol.10, 3934–3944
31. Braun, T., Winter, B., Bober, E. and Arnold, H.H. (1990) Nature 346, 663–665
32. Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988) Cell 55, 989–1003
33. Seharaseyon, J., Bober, E., Hsieh, C., Fodor, W., Franke, U., Arnold, H.H. and Vanin, F. (1990) Genomics 7, 289–293
34. Braun, T., Tannich, E., Buschhausen-Denker, G., and Arnold, H.H. (1989) J.Mol.Cell.Biol.9, 2513–2525 - 30 -
35. Mellentin, J.D., Smith, S.D. and Clearly, M.L. (1989) Cell 58, 77–83
36. Buskin, J.N. and Hauschka, S. (1989) Mol.Cell.Biol.9, 2627–2640
37. Lassar, A.B., Paterson, B.M., and Weintraub, H. (1986) Cell 47, 649–656
38. Pinney, D.F., Pearson-White, S.M., Konieczny, S.F., Latham, K.E. and Emerson, C.P.Jr. (1988) Cell 53, 781–793
39. Thayer, M.J., Tapscott, S.J., Davis, R.L., Wright, W.E., Lassar, A.B. and Weintraub, H. (1989) Cell 58, 241–248
40. Blackwell, T.K. and Weintraub, H. (1990) Science 250, 1104–1110
41. Sassoon, D., Lyons, G., Wright, W.E., Lin, V., Lassar, A., Weintraub, H. and Buckingham, M. (1989) Nature 341, 303–307
42. Ott, M.-O., Bober, E., Lyons, G., Arnold, H.H. and Buckingham, M. (1991) Develop. in press
43. Bober, E., Lyons, G., Braun, T., Cossu, G., Buckingham, M., Arnold, H.H. (1991) J. Cell Biol. in press