Basic helix-loop-helix protein MyoD displays modest DNA binding specificity

Andreas G.E. Künne, Daniel Meierhans, Rudolf K. Allemann*

Department of Chemistry, ETH-Zurich, Universitätstrasse 16, CH-8092 Zurich, Switzerland

Received 4 June 1996

Abstract The expression of MyoD can activate muscle specific genes and myogenic differentiation in many cell types. The hypothesis that the DNA binding specificity of MyoD is responsible for its biological specificity was tested. Homodimers of MyoD bind to E-box containing DNA with high affinity, but do not form stable and well defined complexes with heterologous DNA sequences. The physiologically active heterodimer of MyoD and E12 binds an oligonucleotide containing an E-box sequence with an affinity only two orders of magnitude higher than a completely unrelated DNA sequence, stressing the importance of cooperative interactions with other proteins of the transcriptional machinery for specific gene activation.

Key words: Transcription factor; Protein-DNA interaction; Gene expression; Specificity; Protein-protein interaction

1. Introduction

MyoD (Fig. 1A) is a member of a family of transcription factors involved in the cellular differentiation of the myoblasts of the skeletal musculature [1]. Experiments in transgenic mice have shown that MyoD is responsible for myoblast determination [2,3]. In tissue culture the expression of the MyoD gene is capable of activating previously silent muscle specific genes in myoblasts as well as in many other cell types [4–6]. Surprisingly, a domain of MyoD called the basic helix-loop-helix (BHLH) domain, comprising only 68 amino acid residues, was sufficient for stable myogenic conversion of C3H 10T1/2 fibroblasts suggesting that MyoD acts as a 'master regulator' of muscular differentiation [7]. The X-ray structure analysis of cocrystals of the MyoD BHLH domain and a double stranded oligonucleotide indicated that the HLH domain mediates dimerization of two MyoD monomers, while the DNA is contacted mainly through residues of the basic region [8]. In vivo the functionally active species of MyoD is an heterodimer with either one of the ubiquitously expressed E2A gene products, E12 or E47 [9]. The activity of the MyoD/E12(47) heterodimer as a transcriptional activator depends on the presence of a DNA sequence containing the core motif CANNTG (E-box) [10].

Since an E-box occurs on average every 256 bp in the genome the exquisite physiological specificity of MyoD appears paradoxical. It has been suggested that MyoD displays different affinities for DNA sequences containing different E-boxes. In fact, CASTing [11] and SAAB [12] experiments revealed preferential binding to DNA containing the sequence CAGCTG, and a preference for the nucleotide sequence GA, flanking the E-box on the 5′ side. In addition, the sequence preferences of the MyoD homodimer and of the heterodimer with E47 were different [12] suggesting a second pathway for the generation of sequence specificity. A mechanism in which the specificity of muscular gene expression is solely controlled by the specificity of DNA binding of MyoD, can be functional in vivo only if the affinities of MyoD for individual E-box sequences differ significantly.

We have, therefore, measured the apparent dissociation constants of the complexes of MyoD homodimers and of MyoD/E12 heterodimers with oligonucleotides of different sequences. The apparent dissociation constant (K_D) of the functionally active heterodimer of MyoD and E12 is only two orders of magnitude smaller for the complex with E-box containing DNA sequences than with unrelated DNA. The differences between the affinities for different E-box sequences are even smaller, implying that the specificity of transcriptional activation of muscle specific genes by MyoD cannot be explained by a mechanism which relies solely on the specificity of DNA binding of MyoD. However, most likely it will be achieved through cooperative interaction with other factors of the transcriptional machinery, such as the myocyte specific enhancer factor, MEF-2C [13].

2. Materials and methods

2.1. Expression of MyoD

A fragment of the murine MyoD cDNA (Fig. 1A), coding for the BHLH domain from Ala^{108} to Asp^{166} and with Cys^{133} replaced by Ser (this mutation does not noticeably affect the properties of MyoD [8]) was expressed from the T7 promoter [14] in BL21(DE3) cells essentially as described [15].

2.2. Protein purification

MyoD protein was purified as described for the purification of the BHLH domain of MASH-1 [15]. Protein concentrations were determined by measuring the UV absorptions at 215 and 220 nm [16]. The protein was homogeneous by SDS-gel electrophoresis and by ion exchange HPLC. Laser desorption time of flight mass spectrometry showed sample homogeneity and correct molecular mass. The sequence of the first seven amino-terminal residues was confirmed by automated Edman degradation.

Full-length E12 from rat was expressed and purified as previously described [15].

2.3. Oligonucleotides

Oligonucleotides were purchased from Microsynth. DNA sequences are shown in Fig. 1B. Single stranded oligonucleotides were labelled with [^{32}P]ATP (Amersham) in the presence of T4 polynucleotide ki-

*Corresponding author. Fax: (41) (1) 6321170.
E-mail: ALLEMANN@chem.ethz.ch

Abbreviations: BHLH, basic-helix-loop-helix; CASTing, cyclic amplification and selection of targets; CD, circular dichroism; EMSA, electrophoretic mobility shift assay; HPLC, high performance liquid chromatography; MCK, muscle creatine kinase; SAAB, selected and amplified binding site 'imprints'

PII S0014-5793(96)00707-7

nase (New England Biolabs) and complementary strands annealed by heat denaturation followed by slow cooling to room temperature.

2.4. Electrophoretic mobility shift assays

Aliquots from stock solutions of the proteins were heated to 45°C for 5–10 min in the presence of 10 mM DTT and then allowed to cool to room temperature over a period of 2 h prior to every assay. For experiments with heterodimers equimolar amounts of MyoD and E12 were used. Stock solutions were then serially diluted and incubated in 50 mM Tris (pH 7.9), 6 mM $MgCl_2$, 40 mM ammonium sulfate, 0.2 mM EDTA, 1 mM DTT, and 4% glycerol for 15 min at ambient temperature prior to mixing with the labelled oligonucleotide. These solutions were incubated for 15 min at room temperature and were applied to 4% polyacrylamide gels in 0.9×TAE (pH 7.9), dried, and exposed to Kodak X-OMAT S film at −70°C. Quantitative data were obtained with a Direct Imager (Packard).

2.5. Data analysis

Apparent dissociation constants for the DNA complexes of MyoD were determined as previously described [15]. For the heterodimer between MyoD and E12 the apparent dissociation constant was calculated according to the following formula: $K_D(ME)=\{([E_2D]\times[M])/([MED]\times[E])\}\times K_D(E_2)$.

$K_D(ME)$ and $K_D(E2)$ are the apparent dissociation constants for the DNA complexes of MyoD/E12 hetero- and E12 homodimer, respectively. For $K_D(E2)$ a value of 1.9×10^{-16} M^2 was used [15]. $[E_2D]$, $[M_2D]$, and [MED] can be measured in the titration experiments, while both $[M]=[M]_{tot}-2[M_2D]-[MED]$ and $[E]=[E]_{tot}-2[E_2D]-[MED]$ can be calculated from measured values.

2.6. CD spectroscopy

Spectra were measured on a Jasco J600 circular dichroism spectrophotometer at ambient temperature. The concentration of MyoD was 1 μM in 5 mM Tris (pH 7.0), 0.25 mM DTT.

3. Results and discussion

Results from SAAB and CASTing experiments have been interpreted to show that MyoD could distinguish its DNA target sequences from similar sequences in other regulatory DNA elements through subtle differences of the interactions of the protein with specific and non-specific DNA sequences [11,12]. In addition, based on the strong evidence that the heterodimer between MyoD and the ubiquitous E12/47 protein is the functionally active species in vivo, it has been proposed that heterodimerization might provide an additional source of DNA binding specificity [9].

In order to determine the DNA binding specificity of MyoD, both as a homodimer and as a heterodimer with the E12 protein, we have produced in *E. coli* and purified to apparent homogeneity the BHLH domain of MyoD (Fig. 1A) and characterized its DNA binding properties by electrophoretic mobility shift assays. An oligonucleotide (MCK-S) comprising 17 bp of the IgH enhancer like element of the muscle specific creatine kinase enhancer which contains the central E-box sequence CAGGTG, was used as a DNA probe (Fig. 1B) [17]. Both MyoD and E12 homodimers as well as MyoD/E12 heterodimers formed stable complexes with MCK-S (Fig. 2). Mixing equimolar amounts of MyoD and E12 gave rise to two bands corresponding to the DNA complexes of the E12 homodimer and the heterodimer of E12 and MyoD indicating that they both had similar affinities for MCK-S, while the affinity of the MyoD homodimer for MCK-S was significantly lower (Fig. 2).

Circular dichroism spectroscopy showed that the BHLH domain of MyoD was largely unfolded at a concentration of 1 μM (Fig. 3). Addition of the MCK oligonucleotide to this solution induced a change in the CD spectrum indicative of the transition of MyoD to a largely α-helical conformation upon DNA binding. This transition could be saturated, in that the addition of DNA in excess of a MyoD to DNA-duplex ratio of 2 did not lead to a further change of the CD spectrum, demonstrating that one protein dimer bound to one DNA duplex.

In titration experiments, the apparent dissociation constants (K_D) of the DNA complexes of MyoD with several different oligonucleotides were determined (Table 1). Increasing amounts of protein were added to a constant amount of DNA (Fig. 4A). The concentration of protein $[P]_{1/2}$, at which half of the protein binding sites were filled, was determined from graphs describing the dependence of Φ, the fraction of DNA bound to protein, on the concentration of the unbound

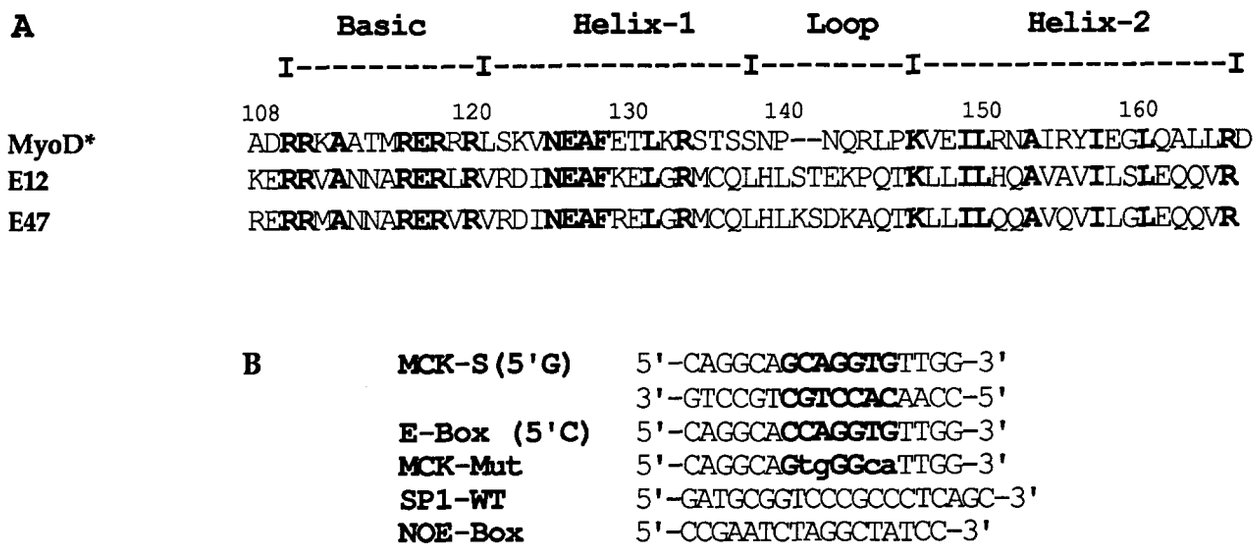

Fig. 1. (A) Sequence alignment of the BHLH domains of murine MyoD* (Cys^{135} of the wild type sequence was replaced with Ser) [8] and E12 and E47 from rat [28]. The numbering system of full-length MyoD is indicated. (B) Sequences of the oligonucleotides used in the EMSAs. The E-box and the nucleotide flanking the E-box on the 5′ side are in bold print.

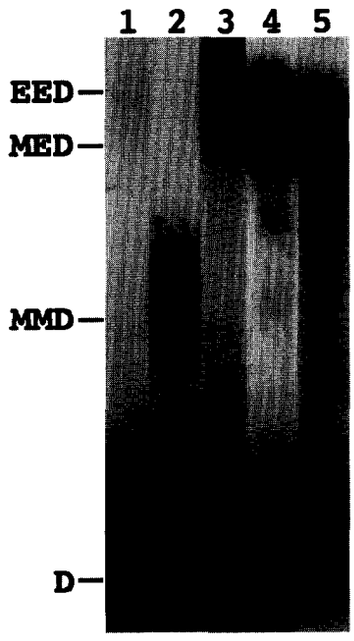

Fig. 2. Binding of MyoD (lane 2) and E12 (lane 3) to the MCK-S oligonucleotide, and of a mixture of equimolar amounts of MyoD and E12 to MCK-S (lane 4) and SP-1 oligonucleotide (lane 5), respectively. Lane 1: MCK-S in the absence of protein.

protein (Fig. 4B) [18]. Since the stoichiometry of the complex between MyoD and DNA is 2:1 (see above) [19], the dissociation constant was calculated as $K_D=([P]_{1/2})^2$. The K_D for the complex of MyoD with MCK-S was found to be 2.2×10^{-14} M^2, in good agreement with a previous report [19]. MyoD homodimer bound E-box containing DNA sequences about two orders of magnitude weaker than E12 homodimer [15] (Table 1).

Titration experiments with DNA sequences not containing E-box sequences (Fig. 1B) were performed in order to determine the DNA binding specificity of the BHLH domain of MyoD. Unlike the oligonucleotides containing E-box sequences, these oligonucleotides did not form well defined complexes with MyoD. However, the great number of observed bands as well as the decrease of the concentration of free DNA with increasing amounts of MyoD (Fig. 4C) indicated the formation of many MyoD-DNA complexes, all with molecular weights smaller than that of the complex of MyoD with E-box containing DNA sequences. The MyoD concentration giving rise to half-maximal DNA binding was more than 10 times higher for heterologous oligonucleotides than for DNA sequences containing an E-box (Fig. 4C).

MyoD folding, homodimerization, and binding of the MyoD dimer to E-box containing DNA sequences are energetically coupled processes (Fig. 3). At high concentration, MyoD forms dimers even in the absence of DNA (K_D ~31 μM) [20]. However, at the submicromolar concentrations where half-maximal DNA binding occurs (Fig. 4B and Table 1), MyoD dimers are dissociated and predominantly exist as unfolded monomers. It is consistent with our data to hypothesize that this DNA induced dimerization is a DNA sequence specific process only induced by DNA containing an E-box sequence. Heterologous DNA sequences appear not to cause dimerization under the conditions of the EMSA. However, due to the great number of positive charges in the basic region of MyoD, these heterologous DNA sequences can, nevertheless, interact with aggregating protein monomers of poorly defined structures, giving rise to a great number of different retarded bands.

While this interpretation of the specificity of DNA binding of MyoD homodimer might be interesting from a structural point of view, it is not relevant to the function of MyoD in vivo, because the physiologically active species is the heterodimer of MyoD and E12 [9]. We therefore investigated the DNA binding properties of the heterodimer. The apparent dissociation constants with several oligonucleotides revealed that MyoD/E12 heterodimers bound E-box containing DNA sequences, with an affinity approximately two orders of magnitude higher than MyoD homodimers (Fig. 2 and Table 1). However, when compared to the homodimer of E12 [15], the apparent dissociation constant was reduced by only threefold.

For E12 homodimers, the biggest difference in K_D for the complexes with E-box containing sequences was observed when the nucleotide flanking the E-box on the 5′ side was changed from a purine to a pyrimidine base [15]. The same observation was made for the heterodimer of MyoD and E12 where the affinity was reduced by one order of magnitude when the guanine on the 5′ side of the E-box of MCK-S was replaced by a cytosine (Table 1). Changing the central two base pairs of the E-box had significantly smaller effects (data not shown).

The crystal structure analysis of MyoD revealed a water mediated contact between the carboxylate of Glu^{118} and N^7 of the adenine flanking the E-box on the 5′ side [8]. This stabilizing interaction will be lost if adenine is replaced by a pyrimidine base. Experiments in transgenic mice have shown that for the myogenin promoter, the proximal E-box, in which a guanine flanks the E-box on the 5′ side, can be replaced with an E-box sequence containing a thymine in the 5′ position without changing the expression pattern of the myogenin gene during embryogenesis [21], thereby providing evidence that a difference in K_D of only one order of magnitude is insignificant in vivo, at least in the context of the myogenin promoter.

The DNA binding specificities of the MyoD/E12 heterodimer and the E12 homodimer were very similar (Table 1). In both cases, the protein concentration required for half max-

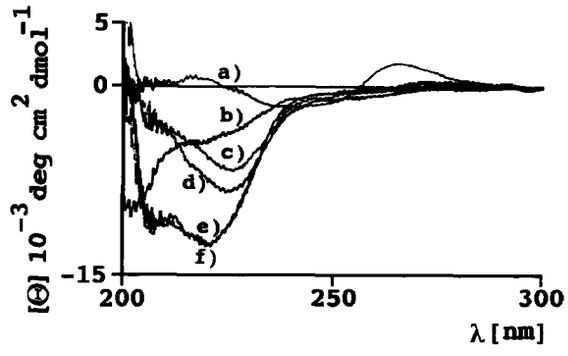

Fig. 3. CD spectra of the BHLH domain of MyoD in the presence of increasing amounts of MCK-S. (a) 0.4 μM MCK-S (no MyoD present); (b–f) [MyoD] = 1 μM with 0, 0.1, 0.3, 0.5, 0.8 μM MCK-S. Spectra b–f are difference spectra between the CD spectrum of the MyoD-MCK-S-complex and MCK-S in the absence of MyoD.

Table 1
Apparent dissociation constants (K_D, M^2) for the DNA complexes of the homodimers of MyoD and E12 [Meierhans], and for the heterodimers of E12 with MyoD and MASH-1 [Meierhans]

	MyoD	E12	MyoD/E12	MASH-1	MASH-1/E12
MCK-S	2.2(±0.5)E14	1.9(±0.9)E16	5.7(±1.0)E17	1.4(±0.1)E14	7.2(±1.2)E17
E-Box 5'C	2.1(±0.1)E14	2.0(±0.4)E15	1.2(±0.5)E15	3.9(±0.4)E14	4.9(±1.3)E16
MCK-MUT	nc	3.6(±1.3)E14	2.0(±1.3)E14	1.4(±0.4)E13	1.3(±0.4)E15
SP1-WT	nc	2.4(±0.8)E14	6.2(±0.8)E15	7.7(±1.2)E14	1.8(±0.4)E15
NOE-BOX	nc	4.4(±0.9)E14	1.9(±1.1)E14	1.4(±0.3)E13	1.2(±0.3)E15

nc: MyoD homodimers do not form stable 2:1 complexes with these oligonucleotides.

imal DNA binding was only 10 times smaller for binding to E-box containing DNA sequences than to heterologous DNA. The heterodimers of MyoD and E12 display, therefore, only marginally higher DNA binding specificity than the heterodimers of E12 with the nerve cell specific protein, MASH-1 (Table 1) [15]. Low DNA sequence specificity seems to be a general feature of the BHLH family of transcription factors. It is noteworthy in this context that removing the leucine zipper from the human transcriptional activator USF, which uses a BHLH-leucine zipper motif for dimerization and sequence specific DNA recognition, leaves a DNA binding region which also shows only marginal specificity of DNA binding [22].

In the light of the crystal structure of the DNA complex of MyoD homodimer [8], the modest DNA binding specificity displayed by BHLH proteins can be rationalized. Most amino acid residues involved in the interaction between MyoD and DNA make sequence independent contacts to the phosphate

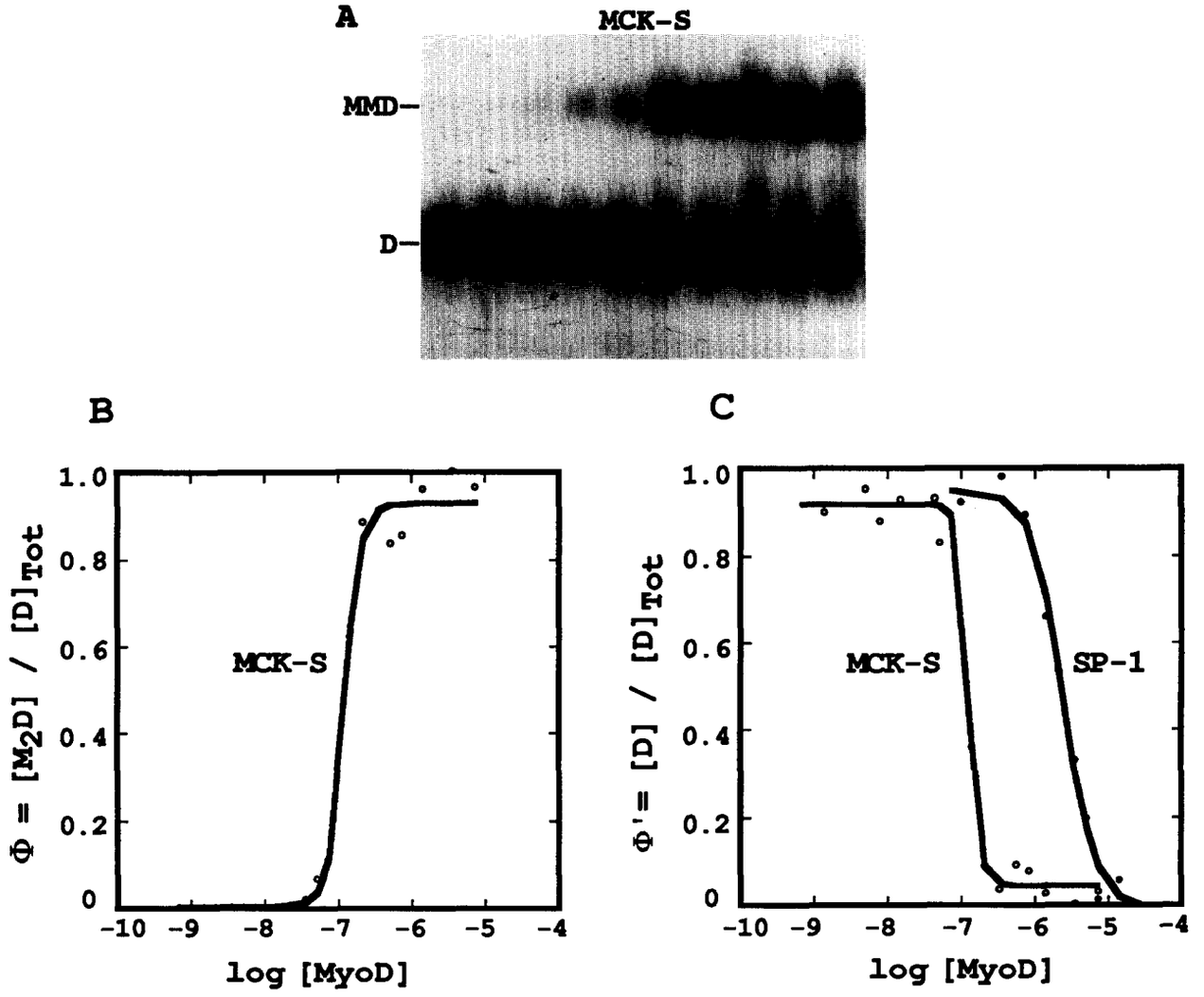

Fig. 4. (A) Autoradiogram of titration of 9.66 nM MCK-S with 14.4, 21.6, 36, 50.5, 72.1, 144.2, 216.2, 360.4, 504.5, and 720.7 nM MyoD. (B) Fraction Φ of bound MCK-S in the MyoD-MCK-S complex vs. the concentration of free MyoD. (C) Fraction Φ′ of free DNA vs. the concentration of the free protein.

backbone. Only two residues of each MyoD monomer make specific contacts to three bases of one DNA half-site. This is in sharp contrast to the large number of specific interactions in the DNA complexes of DNA-methyltransferases which rely on very high sequence specificity for proper function [23,24].

It is clear from our results that sequence specific DNA binding of MyoD cannot explain (at least not as the sole mechanism) the extraordinary specificity of transcriptional activation during muscular differentiation, which must be obtained in large part through cooperative interactions with other transcription factors, such as MEF-2C [13]. Recently published results from transfection experiments with 10T1/2 fibroblasts also indicated that MEF-2C acts as a coregulator to potentiate the myogenic activity of the MyoD/E12 heterodimer through direct interaction between the individual DNA binding domains [25].

Both the affinity and the specificity of DNA binding of the MyoD/E12 heterodimer are comparable to that of the E12 homodimer and the question arises why E12 does not induce muscular differentiation in the absence of MyoD. The replacement of only two amino acid residues in the basic region and of one residue in the amino terminal region of helix 1 of E12 with the corresponding amino acids from MyoD are enough for the acquisition of myogenic specificity of E12 [26]. The two residues of the basic region are located on the interface between the DNA duplex and the protein, suggesting that they might be involved in controlling the specificity of DNA binding, while the third residue required for myogenic conversion protrudes into the solvent [8,27]. Our data show that these three amino acid changes do not significantly alter the DNA binding specificity of E12. Rather, these three residues control the potential of E12 and MyoD to interact with other proteins involved in the transcriptional regulation of muscle specific genes. In such a mechanism the basic region is responsible for sequestering the BHLH proteins to DNA, while sequence specificity and, consequently, the specificity of transcriptional activation is mainly achieved through cooperative interactions.

Acknowledgements: This work was supported by an ETH grant (R.K.A.). We thank Dr. Carl O. Pabo for providing the MyoD plasmid; David J. Anderson for providing the cDNA of E12; Lesley A. Tannahill and Martin Sieber for critical reading of the manuscript and for helpful discussions.

References

[1] Sassoon, D., Lyons, G., Wright, W.E., Lin, V., Lassar, A., Weintraub, H. and Buckingham, M. (1989) Nature 341, 303–307.
[2] Hasty, P., Bradley, A., Morris, J.H., Edmondson, D.G., Venuti, J.M., Olson, E.N. and Klein, W.H. (1993) Nature 364, 501–506.
[3] Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S., Nanoka, I. and Nabeshima, Y. (1993) Nature 364, 532–535.
[4] Buckingham (1992) Trends Genet. 8, 144–149.
[5] Olson, E. (1990) Genes Dev. 4, 1454–1461.
[6] Weintraub, H., Davis, R., Tapscott, S., Thayer, M, Krause, M., Benezra, R., Blackwell, T.K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y. and Lassar, A. (1991) Science 251, 761–766.
[7] Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.-F., Weintraub, H. and Lassar, A.B. (1988) Science 242, 405–411.
[8] Ma, P.C.M., Rould, M.A., Weintraub, H. and Pabo, C.O. (1994) Cell 77, 451- 459.
[9] Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronona, A., Baltimore, D. and Weintraub, H. (1991) Cell 66, 305–315.
[10] Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D. and Weintraub, H. (1989) Cell 58, 823–831.
[11] Wright, W.E., Binder, M. and Funk, W. (1991) Mol. Cell. Biol. 11, 4104–4110.
[12] Blackwell, T.K. and Weintraub, H. (1990) Science 250, 1104–1110.
[13] Olson, E.N., Perry, M. and Schultz, R.A. (1995) Dev. Biol. 172, 2–14.
[14] Studier, F.W. (1991) J. Mol. Biol. 219, 37–44.
[15] Meierhans, D., el Ariss, C., Neuenschwander, M., Sieber, M., Stackhouse, J.F. and Allemann, R.K. (1995) Biochemistry 34, 11026–11036.
[16] Wetlaufer, D.R. (1962) Adv. Protein Chem. 17, 303–390.
[17] Buskin, J.N. and Hauschka, S.D. (1989) Mol. Cell. Biol 9, 2627–2640.
[18] Clore, G.M., Gronenborn, A.M. and Davies, R.W. (1982) J. Mol. Biol. 155, 447–466.
[19] Sun, X.-H. and Baltimore, D. (1991) Cell 64, 459–470.
[20] Fairman, R., Beran-Steed, R.K., Anthony-Cahill, S.J., Lear, J.D., Stafford III, W.F., DeGardo, W.F., Benfield, P.A. and Brenner, S.L. (1993) Proc. Natl. Acad. Sci USA 90, 10429–10433.
[21] Yee, S.-P. and Rigby, P.W.J. (1993) Genes Dev. 7, 1277–1289.
[22] Sha, M., Ferré-D'Amaré, A.R., Burley, S.K. and Goss, D.J. (1995) J. Biol. Chem. 270, 19325–19329.
[23] Reinisch, K.M., Chen, L., Verdine, G.L. and Lipscomb, W.N. (1995) Cell 82, 143–153.
[24] Klimasaukas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) Cell 76, 357–369.
[25] Molkentin, J.D., Black, B.L., Martin, J.F. and Olson, E.N (1995) Cell 83, 1125–1136.
[26] Davis, R.L. and Weintraub, H. (1992) Science 256, 1027–1030.
[27] Ellenberger, T., Fass, D., Arnaud, M. and Harrison, S.C. (1994) Genes Dev. 8, 970–980.
[28] German, M.S., Blanar, M.A., Nelson, C., Moss, L.G. and Rutter, W.J. (1991) Mol. Endocrinol. 5, 292–299.