

Establishment of Distinct MyoD, E2A, and Twist DNA Binding Specificities by Different Basic Region-DNA Conformations

THIPHAPHONE KOPHENGNAVONG, JENNIFER E. MICHNOWICZ, AND T. KEITH BLACKWELL*

Center for Blood Research and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115

Received 23 July 1999/Returned for modification 1 September 1999/Accepted 23 September 1999

Basic helix-loop-helix (bHLH) proteins perform a wide variety of biological functions. Most bHLH proteins recognize the consensus DNA sequence CAN NTG (the E-box consensus sequence is underlined) but acquire further functional specificity by preferring distinct internal and flanking bases. In addition, induction of myogenesis by MyoD-related bHLH proteins depends on myogenic basic region (BR) and BR-HLH junction residues that are not essential for binding to a muscle-specific site, implying that their BRs may be involved in other critical interactions. We have investigated whether the myogenic residues influence DNA sequence recognition and how MyoD, Twist, and their E2A partner proteins prefer distinct CAN NTG sites. In MyoD, the myogenic BR residues establish specificity for particular CAN NTG sites indirectly, by influencing the conformation through which the BR helix binds DNA. An analysis of DNA binding by BR and junction mutants suggests that an appropriate BR-DNA conformation is necessary but not sufficient for myogenesis, supporting the model that additional interactions with this region are important. The sequence specificities of E2A and Twist proteins require the corresponding BR residues. In addition, mechanisms that position the BR allow E2A to prefer distinct half-sites as a heterodimer with MyoD or Twist, indicating that the E2A BR can be directed toward different targets by dimerization with different partners. Our findings indicate that E2A and its partner bHLH proteins bind to CAN NTG sites by adopting particular preferred BR-DNA conformations, from which they derive differences in sequence recognition that can be important for functional specificity.

A large family of transcriptional regulators is defined by the basic helix-loop-helix (bHLH) motif (40), in which a DNA-binding basic region (BR) lies immediately amino terminal to the HLH dimerization segment (17, 41, 55). In metazoans, bHLH proteins are involved in specification of multiple cell types (33, 43, 56). Some bHLH family members function as homodimers, but others appear to act together with a heterodimeric partner (56). For example, the closely related bHLH proteins that mediate myogenic differentiation, including MyoD, are thought to function as heterodimers with E proteins, a widely expressed bHLH protein subgroup that is exemplified by the E2A proteins (14, 17, 32, 42). Most bHLH protein dimers bind to the consensus CAN NTG (the E box; the consensus sequence is underlined throughout the text) with each respective BR binding to a half site (9, 19–21, 35, 45, 48). Given the many regulatory processes in which bHLH proteins are involved, the apparent simplicity of the CAN NTG consensus raises the important question of how different bHLH proteins act only on appropriate target genes (56).

In part, the specificity with which bHLH proteins function derives from preferential recognition of different classes of CAN NTG sites by different bHLH protein subgroups. The HLH segment consists of a parallel, left-handed, four-helix bundle (Fig. 1) (19–21, 35, 45, 48). The BR is unstructured in solution (2) but when bound to DNA, it extends N terminally from the HLH segment as an α helix that crosses the major groove (Fig. 1). Crystallographic analyses have revealed some differences in how these proteins bind DNA. For example, in Myc family and related bHLH proteins, an arginine (Arg) residue at BR position 13 (Fig. 2) specifies recognition of CACGTG sites (7, 16, 25, 54) by contacting bases in the center

(20, 21, 48). However, it still is not understood how bHLH proteins which have a different amino acid at BR position 13 (Fig. 2) bind preferentially to distinct CAN NTG sites (9, 16) or how bHLH proteins establish differences in flanking sequence selectivity (9, 23, 24) that can be of biological importance (1, 30).

Many bHLH proteins that lack R₁₃, including MyoD and other E2A partners (Fig. 2), can bind to similar DNA sequences *in vitro* but act on different tissue-specific genes (56). Cooperative or inhibitory relationships with other transcriptional regulators might contribute to this specificity (34, 39, 46, 58), but it is not likely to derive entirely from other lineage-specific factors, because MyoD can induce myogenesis in many different cell types (56). Initiation of myogenesis by MyoD and other myogenic bHLH proteins depends on three residues that are located within the BR and the BR-HLH junction (A₅, T₆, and K₁₅ [Fig. 1 and 2]). These residues, which we refer to as myogenic are not essential for binding a muscle-specific site *in vitro* or *in vivo*, suggesting that they are involved in other critical interactions (11, 17, 18, 47, 57). These interactions have been proposed to involve distinct cofactors (11, 17, 57) and the unmasking of an activation domain in MyoD or the myogenic cofactor MEF2 (3, 5, 29, 57). In the MyoD-DNA structure, K₁₅ is oriented away from the DNA, but A₅ and T₆ face the major groove and could not contact other proteins directly (35) (Fig. 1). However, the latter two residues could influence protein-protein interactions indirectly, by affecting how the BR helix is positioned on the DNA (35). Although substitutions at these positions might not substantially impair binding to particular CAN NTG sites, it is important to determine whether they might have more subtle influences on sequence specificity that could reflect conformational effects.

We have determined that the myogenic residues A₅ and T₆ establish the characteristic MyoD sequence preference, which includes a CAGCTG core. Individual substitutions at these BR positions simultaneously alter preferences for multiple bases

* Corresponding author. Mailing address: Center for Blood Research and Department of Pathology, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115. Phone: (617) 278-3150. Fax: (617) 278-3131. E-mail: blackwell@cbr.med.harvard.edu.

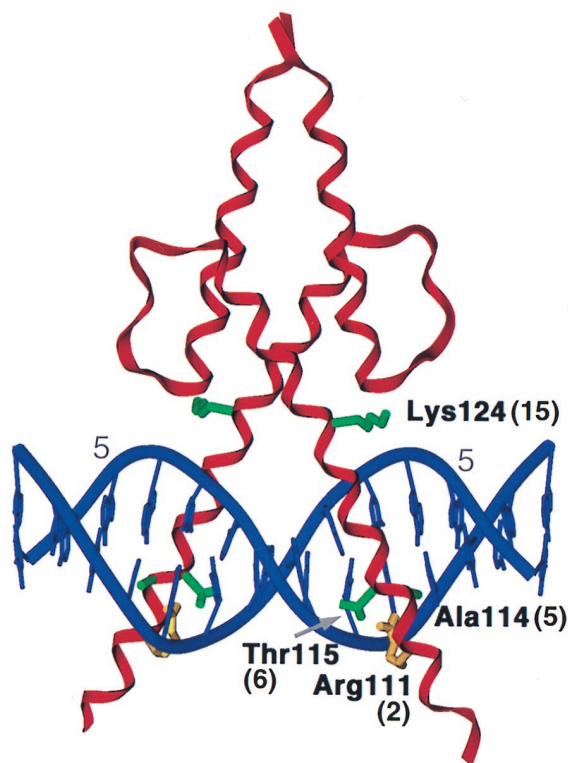


FIG. 1. A MyoD-DNA complex. In this X-ray crystallographic structure (35), a MyoD homodimer is bound to the sequence AACAGCTGTT, which corresponds to its preferred recognition consensus (9). Residues are numbered as in full-length MyoD, and their positions as specified in Fig. 2 and the text are indicated in parentheses. Binding site positions ± 5 (numbered as in Fig. 3A) are indicated by grey numerals. Side chains are shown only for the myogenic residues (green) (18) and Arg 111 (R_2) (gold).

that MyoD does not contact directly (35), indicating that these preferences are determined indirectly, by how the BR helix is positioned on the DNA. This mechanism is distinct from the standard model for sequence specificity, in which preferred bases are contacted directly (44, 50). The corresponding BR residues are also required for the sequence preferences of E2A proteins, which can recognize either of two distinct half-sites depending on their dimerization partner. E2A homodimers and E2A-MyoD heterodimers bind to asymmetric sites that include a CACCTG core. In contrast, as a heterodimer with the bHLH protein Twist, E2A binds preferentially to half of the symmetric sequence CATATG. The preference of E2A for the former asymmetric sites depends not only on the BR sequence but also on BR positioning that involves the junction region. An analysis of DNA binding by MyoD and E2A junction and BR mutants indicates that a MyoD-like sequence specificity is associated with, but not sufficient for, myogenesis. This supports the model that the BR-junction region is also involved in other critical interactions. The results suggest that E2A and its partner bHLH proteins bind DNA by adopting a limited number of preferred BR conformations, each of which is associated with a characteristic DNA sequence preference. They also indicate that binding of cofactors to the MyoD BR might be influenced by how it is positioned on the DNA and are consistent with the idea that relatively subtle differences in binding sequence recognition can modulate bHLH protein activity in vivo.

MATERIALS AND METHODS

Mutagenesis, protein expression, and DNA binding assays. The various MyoD and E2A mutants used in this study have been described previously (17, 18, 57), with the exception of the MyoD mutants shown in Fig. 8A. For construction of those mutants, a *SalI* site that did not alter the encoded amino acid sequence was created at MyoD BR positions 10 and 11 (Fig. 2) by PCR. BR mutants were then generated by PCR using *Pfu* or Vent polymerase and introduced into this MyoD (*SalI*) construct as *PmlI-SalI* fragments. Junction region mutations were created similarly by PCR and inserted into MyoD (*SalI*) as *SalI-NarI* fragments. Constructs with both BR and junction mutations were produced by introduction of a mutant *PmlI-SalI* or *SalI-NarI* fragment into the appropriate BR or junction mutant construct. All of these mutations were confirmed by DNA sequencing.

For the in vitro selection experiment shown in Fig. 3, full-length MyoD was expressed in bacteria from a pRK171a-based construct (pT7-MyoD) described previously (53). The MD(E12B), MD(E12B-A), and MD(E12B-AT) mutations (57) were each introduced into this construct within a *PmlI-MluI* fragment. These proteins were expressed by isopropyl- β -D-thiogalactopyranoside induction in *Escherichia coli* BL21(DE3)/pLysS cells as described previously (51) and then purified to >90% homogeneity by precipitation in 0.6 M $(\text{NH}_4)_2\text{SO}_4$. Precipitated protein was resuspended in a mixture containing 10% glycerol, 20 mM HEPES (pH 7.6), 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 1 μg of leupeptin per ml, and 1 μg of pepstatin per ml.

Other proteins were expressed by in vitro translation (Promega), with in vitro transcription and translation performed in separate steps. Protein expression was carefully quantitated by ^{35}S -labeled translation and sodium dodecyl sulfate-polyacrylamide gel electrophoresis. These procedures and those for electrophoretic mobility shift assay (EMSA) have been described previously (31). Each EMSA was performed at room temperature and analyzed by autoradiography or phosphorimaging. Individual oligonucleotide sites were 21 bp in length and differed from the MyoD consensus oligonucleotide (7) only at the positions indicated. The MCK-R site corresponds to the right E-box site in the muscle creatine kinase enhancer (18).

In vitro selection experiments. Populations of preferred binding sites were isolated by sequential in vitro selection and PCR amplification essentially as described previously (6, 7, 9). During each selection round, DNA that was bound by the protein complex of interest was isolated by EMSA and then amplified by PCR for the next round. In each EMSA selection, care was taken to ensure that sufficient quantities of labeled bound DNA were recovered to maintain a representative population of sequences. These experiments were initiated with 0.5 ng of ^{32}P -end-labeled starting library. In each subsequent selection round, selections were performed with approximately 0.1 ng of amplified ^{32}P -body-labeled DNA. In selections for binding to partially purified bacterially expressed MyoD mutant proteins (Fig. 3), these protein preparations were not quantitated, but instead sequential dilutions of these samples were tested for binding. Bound sequences were then recovered and amplified from a sample in which less than 10% of the input DNA was in the bound fraction. This strategy ensured selection of optimal binding sequences. The final selected binding site pool was sequenced directly, using a ^{32}P -end-labeled primer as described previously (7). Mouse proteins were used in these selections with the exception of Twist, which was from *Xenopus*. Binding site competition analyses (not shown) demonstrated that its binding preferences were indistinguishable from those of mouse Twist, which was used in the EMSA analyses shown.

RESULTS

Myogenic BR residues and MyoD DNA binding preferences. Identification of the myogenic BR residues stemmed originally from studies in which the MyoD BR was replaced with that of E12, a product of the alternatively spliced E2A gene (40). This MyoD mutant [MD(E12B) (Fig. 2)] binds to a muscle-specific regulatory site as a heterodimer with E2A proteins either in vitro or in vivo, but it cannot induce myogenesis in a cell culture assay or activate transcription through a muscle-specific enhancer (17, 57). Resubstitution of the myogenic residues A_5 and T_6 (Fig. 2) in MD(E12B) restores its activity in these functional assays (57). Similar results are obtained when A_5 and T_6 are mutated within MyoD (18, 29, 57) and when analogous substitutions are made in the context of the myogenic bHLH protein myogenin (11). These experiments implicate A_5 and T_6 in mechanisms that are of functional importance but not essential for binding to a particular muscle-specific DNA sequence.

We used an in vitro selection strategy (9) to test whether such mutations might have more subtle effects on how MyoD binds specifically to DNA. To identify sequences to which

			<u>MUSCLE:</u>
<u>BASIC REGIONS:</u>			
MyoD	K R K T T N A D <u>R R K A A T M R E R R R</u> L S K V		++++
E12	Q K A E R E K E <u>R R V A N N A R E R L R</u> V R D I		No
Twist	Q S Y E E L Q T <u>Q R V M A N V R E R Q R</u> T Q S L		ND
<u>MUTANTS:</u>			
MD (E12B)	[Q K A E R E K E <u>R R V A N N A R E R L R</u>] L S K V		No
MD (E12B-A)	[Q K A E R E K E <u>R R V A A N A R E R L R</u>] L S K V		No
MD (E12B-AT)	[Q K A E R E K E <u>R R V A A T A R E R L R</u>] L S K V		+
MD (E12BJ)	[Q K A E R E K E <u>R R V A N N A R E R L R</u>] V R D I		No
E12 (MDB)	Q K A [T T N A D R R K A A T M R E R R R] V R D I		No
E12 (MDBJ)	Q K A [T T N A D R R K A A T M R E R R R] L S K V		++
E12 (AT, MDJ)	Q K A E R E K E <u>R R V A A T A R E R L R</u> L S K V		+
E12 (AT, K)	Q K A E R E K E <u>R R V A A T A R E R L R</u> V R K I		+
E12 (AT)	Q K A E R E K E <u>R R V A A T A R E R L R</u> V R D I		No
	1 5 10 15		
	Basic	Junction	

FIG. 2. Myogenic activity of MyoD and E12 BR and junction mutants. Each of these mutants has been described previously (18, 57), and their sequences are compared with sequences from mouse MyoD, E12, and Twist. Amino acids that are identical to those of MyoD are underlined, positions that are conserved in most bHLH proteins are shaded, and entire BR and junction regions that have been swapped are bracketed. The column at the right indicates the relative activities of these proteins when assayed previously by transfection for conversion of cultured cells into muscle (18, 57); activity is denoted as ++++ (frequency of myogenic conversion obtained with wild-type MyoD), ++ (30 to 50% of that obtained with MyoD), + (5 to 30% of that obtained with wild-type MyoD), No (myogenic conversion was not detected), or ND (not done).

these mutants bind preferentially, we used sequence libraries in which only positions within and flanking the CAN NTG consensus are randomized (Fig. 3A), so that the position of bHLH protein binding along the DNA is fixed. This strategy makes it possible to sequence the selected sites as a pool and thereby to analyze a very large population of selected sites simultaneously (8, 9). It reveals the relative preferences for individual bases at each site position and can detect subtle differences that might not be identified through more conventional approaches.

This assay has previously shown that the preferred MyoD binding consensus is (G/A) ACAGCTG(T/C) (Fig. 3B and C) and that the E2A proteins E12 and E47 overlap considerably with MyoD in their binding properties but prefer sites that have an asymmetric CACCTG core sequence (Fig. 3C) (9). However, in contrast to either of these proteins, the MD(E12B) mutant prefers the sequence (G/A) CCATATGG(T/C), which differs from the MyoD preferred site over the eight central base pairs and contains the distinct core sequence CAT ATG (Fig. 3B and C). This sequence and related elements are normally targeted by the bHLH protein Twist, an E-protein partner that is involved in mesodermal cell fate specification (15, 27, 37, 52, 60) (Fig. 2). Back-substitution of A₅ of MyoD into MD(E12B), which is not sufficient for myogenic activity in cell culture assays (57), results in preferences that are slightly more similar to those of MyoD at positions ±4 [MD(E12B-A) (Fig. 2, 3B, and 3C)]. However, introduction of

both A₅ and T₆, which restores myogenesis (11, 57), results in preferences across the entire site that are indistinguishable from those of MyoD [MD(E12B-AT) (Fig. 2, 3B, and 3C)].

To determine whether these sequence preferences reflect significant differences in binding affinity and specificity, we compared levels of binding of these proteins to individual oligonucleotides that correspond to the MyoD and Twist preferences and differ only at positions within and adjacent to the CAN NTG consensus (Fig. 3D). Supporting the in vitro selection findings, both MyoD and MyoD(E12B-AT) homodimers bound with higher affinity to the preferred MyoD site than to the Twist site (Fig. 3D, lanes 1, 4, 5, and 8). In contrast, the Twist site was preferred by MD(E12B) and, to a lesser extent, MyoD(E12B-A) (Fig. 3D, lanes 2, 3, 6, and 7). In a binding competition assay, specific DNA binding by MD(E12B-AT) was competed much more effectively by the MyoD site (Fig. 4A, lanes 4, 7, 10, 13, and 16), and binding by either MD(E12B) or MD(E12B-A) was competed better by the Twist site (Fig. 4B, lanes 2, 3, 8, 9, 14, and 15). A c-Myc preferred site (CACGTG [not shown]) was a relatively poor competitor of binding by each of these proteins (Fig. 4A and B, lanes 17 to 19). The data show that introduction of A₅ and T₆ into MD(E12B) restores not only myogenic activity (Fig. 2) but also the MyoD DNA binding preference. This substitution affects sequence recognition across 4 bp within each half-site (Fig. 3A and B), indicating a global effect on how the BR helix is positioned on the DNA. The finding that MD(E12B) is distinct

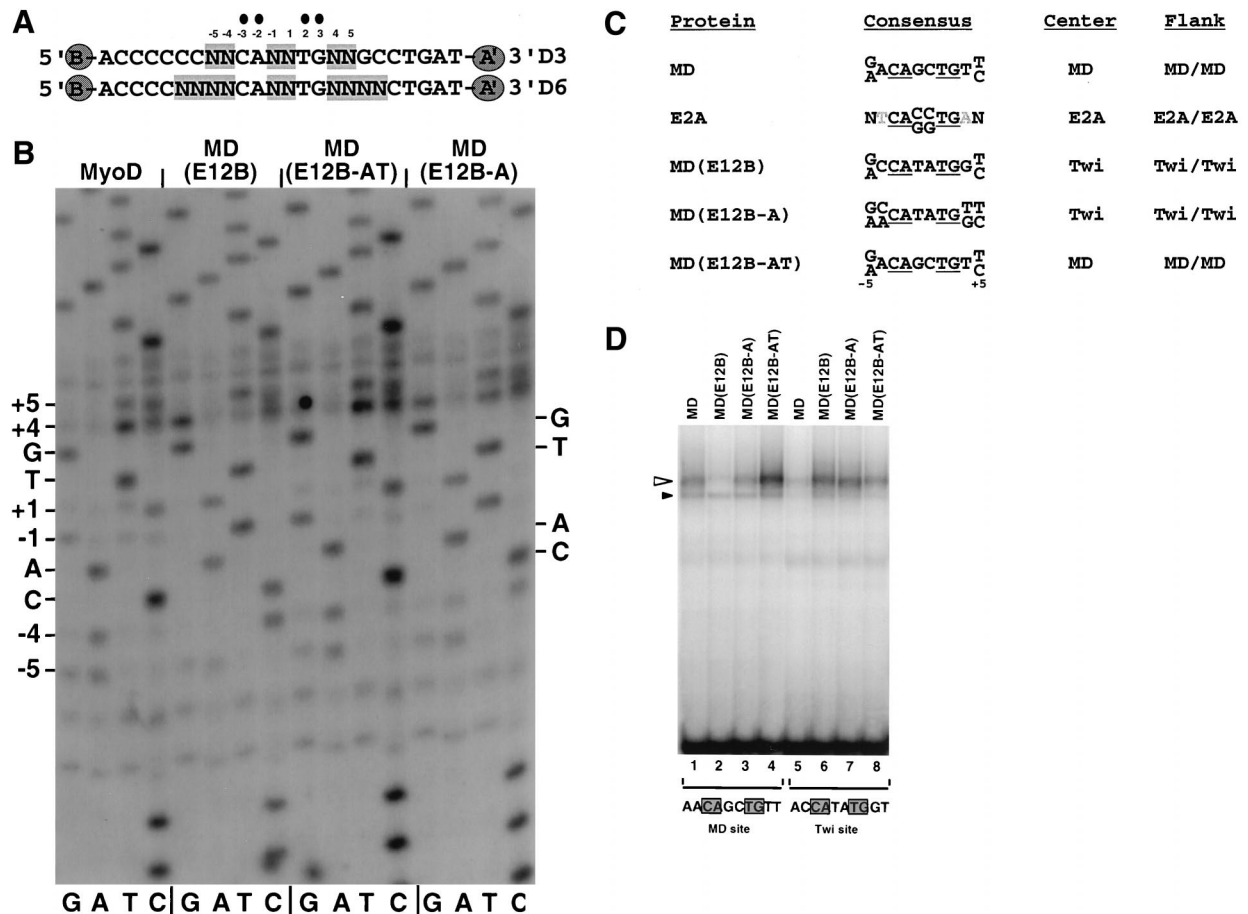


FIG. 3. In vitro selection assay of binding site preferences. (A) Core sequences of the random sequence oligonucleotide libraries D3 and D6 (8, 9). In each library, the bases shown are flanked by sequences which correspond to primers (A and B) that allow selected sequences to be recovered by PCR. A' indicates that primer A corresponds to the opposite strand. (B) Sequences of preferred binding sites. Starting with the D6 oligonucleotide random sequence library (A), three rounds of sequential selection and PCR amplification were performed for binding to the proteins indicated. A sample of the final selected population of binding sites was then sequenced directly as a pool and analyzed by autoradiography. The MyoD preferences at positions ± 1 described previously (9) are more prominent after additional selection rounds (not shown). (C) Summary of sequence preferences identified by in vitro selection in panel B. MyoD and E2A homodimer preferences were described in reference 9. Binding site positions are numbered as in panel B, and grey letters indicate bases that were selected against. The CAN NTG consensus that was fixed in these experiments is underlined. Twist, Twist. (D) Binding of MyoD BR mutants to individual oligonucleotide sites, which differed only at the sequences shown. In this EMSA, which was analyzed by phosphorimaging, each sample contained the indicated in vitro-translated protein at a concentration of 40 pM and DNA that was labeled to the same specific activity at 550 pM. Specific and background species are indicated by open and closed triangles, respectively.

from either MyoD or E12 in its binding sequence preference also indicates that DNA recognition by an E2A BR can be profoundly influenced by its molecular context.

Influence of BR positioning on MyoD-E2A and Twist-E2A heterodimer sequence preferences. Twist and E2A proteins appear to cooperate in vivo to regulate transcription through CAT ATG sites (27), suggesting that the DNA sequence recognition properties of E2A might be altered by heterodimerization with Twist. However, an alternative possibility is that functional Twist-E2A recognition sites are distinct from their in vitro binding preference (28). To address this question, we performed in vitro selection on Twist-E12 complexes. Twist homodimers and Twist-E12 heterodimers both preferred sites that contain the core sequence CAT ATG (Fig. 5A and B). They were similar to MD(E12B) and especially to MD(E12B-A) in their preferences at ± 4 but selected MyoD-like sequences at ± 5 (Fig. 3B and 3C, 5A, and 5B). The symmetry of this preferred sequence suggests that in the Twist-E12 protein-DNA complex, the Twist and E12 BRs each prefer the same half-site sequence. In contrast, and as observed previously (9), MyoD-E12 heterodimers selected a MyoD-like half

site at positions +4 and +5, an E2A-like half-site at -4 and -5, and CC or GG bases in the center of the site (Fig. 5A and B), indicating asymmetric binding. Apparently, an E2A BR normally prefers distinct half-sites in the context of these two bHLH dimerization partners, indicating an intermolecular effect on how it interacts specifically with DNA.

To investigate how heterodimer formation influences the binding preferences of the E12 and MyoD BRs, we performed in vitro selection on combinations of MyoD and E12 BR mutants. When the BR of one partner within a MyoD-E12 heterodimer was substituted with that of the other, the heterodimer binding preferences outside the CAN NTG consensus corresponded to those of the individual BRs. For example, unlike MD(E12B) homodimers (Fig. 3B and C), heterodimers of MD(E12B)+E12 preferred wild-type heterodimer sequences in the center of the site, and selected E2A-like sequences in both flanking regions, at ± 4 and ± 5 (Fig. 5A and B). A heterodimer of MyoD and an E12 protein containing the MyoD BR [E12(MDB) (Fig. 2A)] similarly selected a wild-type heterodimer preference within the CAN NTG motif but preferred a MyoD-like sequence at ± 4 and ± 5

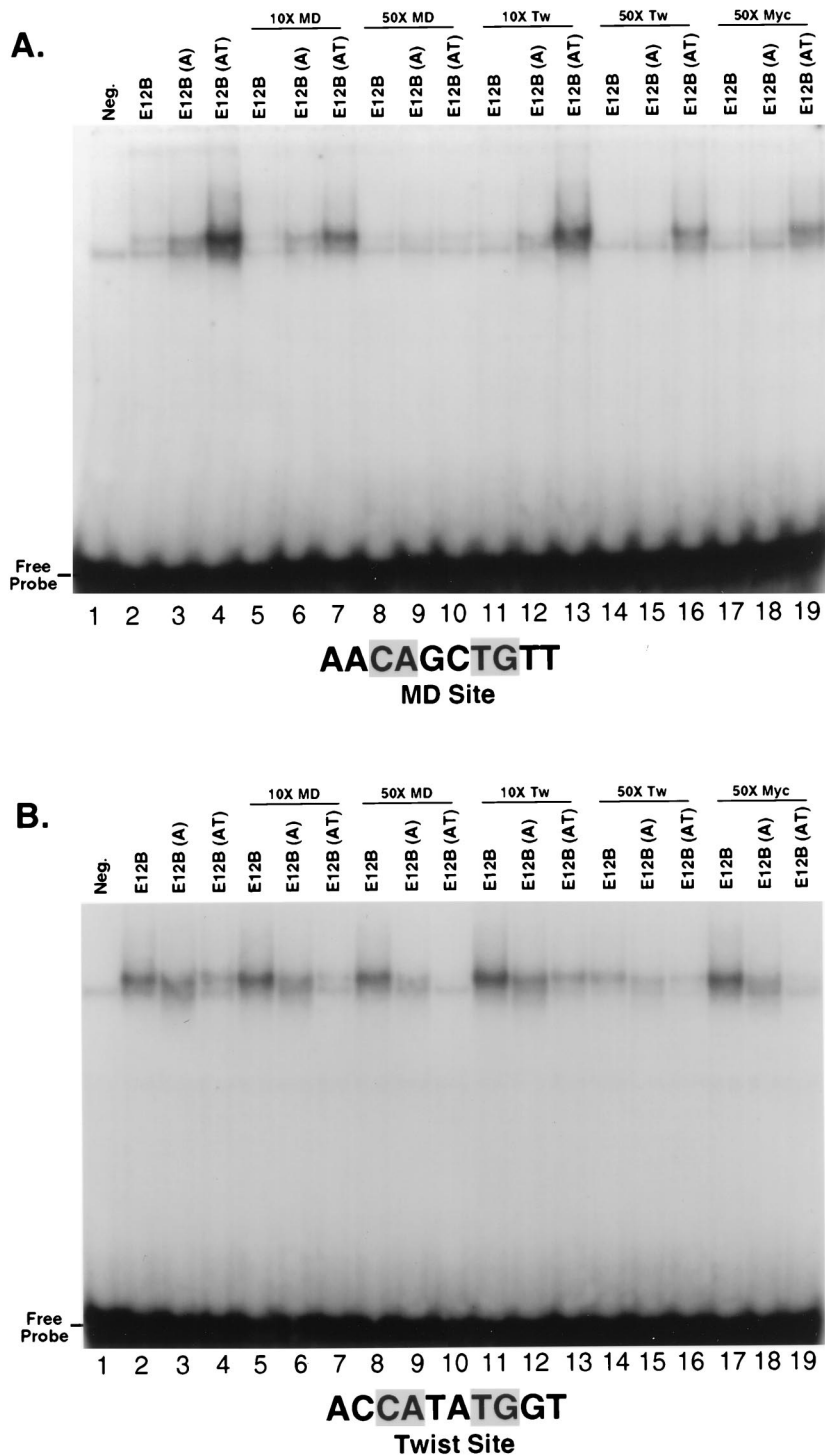


FIG. 4. Specificity of MyoD BR mutant DNA binding. (A) Competition analysis of binding to the labeled MyoD preferred site, analyzed by EMSA and autoradiography. The indicated in vitro-translated proteins and DNA labeled to the same specific activity were present at concentrations of 50 and 900 pM, respectively. When the samples were mixed, unlabeled competitor DNA sites were added at the indicated ratios relative to the labeled probe. Tw, Twist. (B) Competition analysis of binding to the Twist preferred site, performed as for panel A.

(Fig. 5A and B). In contrast, MD(E12B)-E12(MDB) heterodimers had a binding preference more similar to that of Twist (Fig. 5A and B), indicating that placement of each BR in the protein context of the other partner affected binding over the entire site. A striking aspect of our findings is that each of

the mutant homo- or heterodimer protein complexes that we have examined selected sequences that correspond to particular patterns preferred by MyoD, E2A, or Twist protein (Fig. 3C and 5B).

These in vitro selection findings were supported by assays of

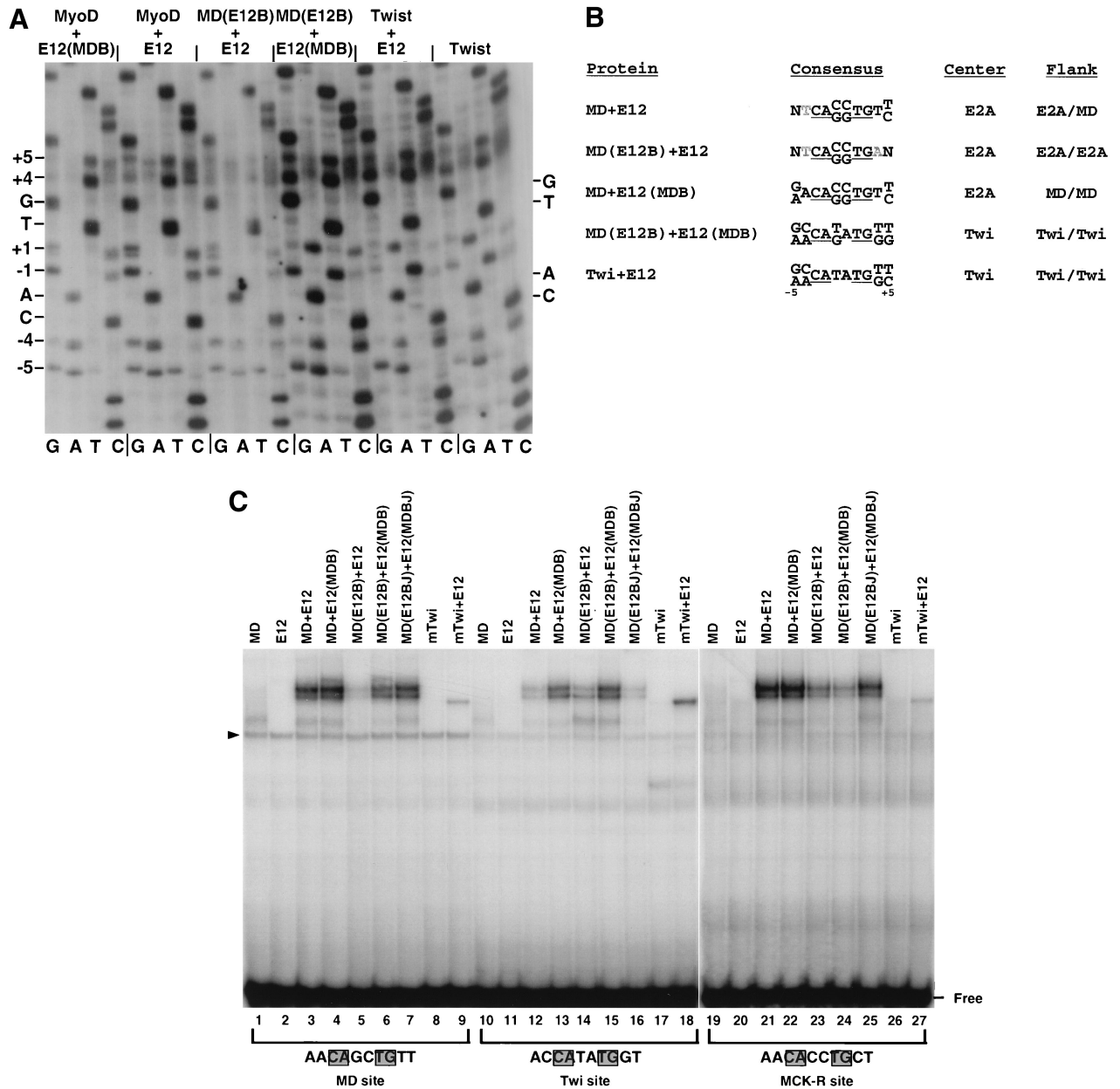


FIG. 5. Binding site preferences of MyoD, E2A, and Twist heterodimer complexes. (A) In vitro selection analysis of binding site preferences. Four rounds of selection from the D3 library (Fig. 3A) were performed as for each in vitro-translated protein complex. In each case, the heterodimer complex could be easily identified in the EMSA on the basis of mobility (9), particularly because E12 homodimers bind DNA poorly (Fig. 9). In the Twist homodimer selection, binding to Twist-E12 heterodimers was selected for in the first round, because of the relatively low level of Twist homodimer binding. Subsequent rounds were performed with Twist homodimers. Each sample was analyzed by sequencing and autoradiography as for Fig. 3B. (B) Summary of sequence preferences identified in panel A, depicted as in Fig. 3C. MyoD-E2A heterodimer preferences were also described previously (9). Twi, Twist. (C) Binding of bHLH heterodimers to individual preferred sites, analyzed by EMSA and phosphorimaging. E2A-derived proteins were present at a concentration of 8 pM, and Twist and MyoD-derived proteins were present at 19 pM. The indicated DNA sites that had been labeled to the same specific activity were present at 550 pM. The MCK-R site differs from the others only at the positions shown. A background species is indicated by a triangle.

binding to individual sites, including a sequence from a muscle-specific regulatory region (MCK-R). This site, which corresponds to the MyoD-E12 heterodimer in vitro binding preference and responds to MyoD in vivo, was used in the original analysis of the myogenic residues (9, 17, 57). In an EMSA, MyoD-E12 heterodimers bound with higher affinity to either the MCK-R or MyoD site than to the Twist site (Fig. 5C, lanes 3, 12, and 21). MyoD(E12B)-E12 heterodimers only slightly preferred the MCK-R heterodimer site to the Twist site but

appeared to prefer either of these sequences to the MyoD site (Fig. 5C, lanes 5, 14, and 23). As the preferences of MD(E12B-A) and MD(E12B-AT) homodimers would predict, introduction of both A₅ and T₆ into MD(E12B) altered its sequence preferences as a heterodimer with E12, so that they were more similar to those of MyoD (not shown). MyoD-E12(MDB) heterodimers only modestly preferred the MyoD or MCK-R site in comparison to the Twist site (Fig. 5C, lanes 4, 13, and 22). In contrast, the Twist site was preferred by

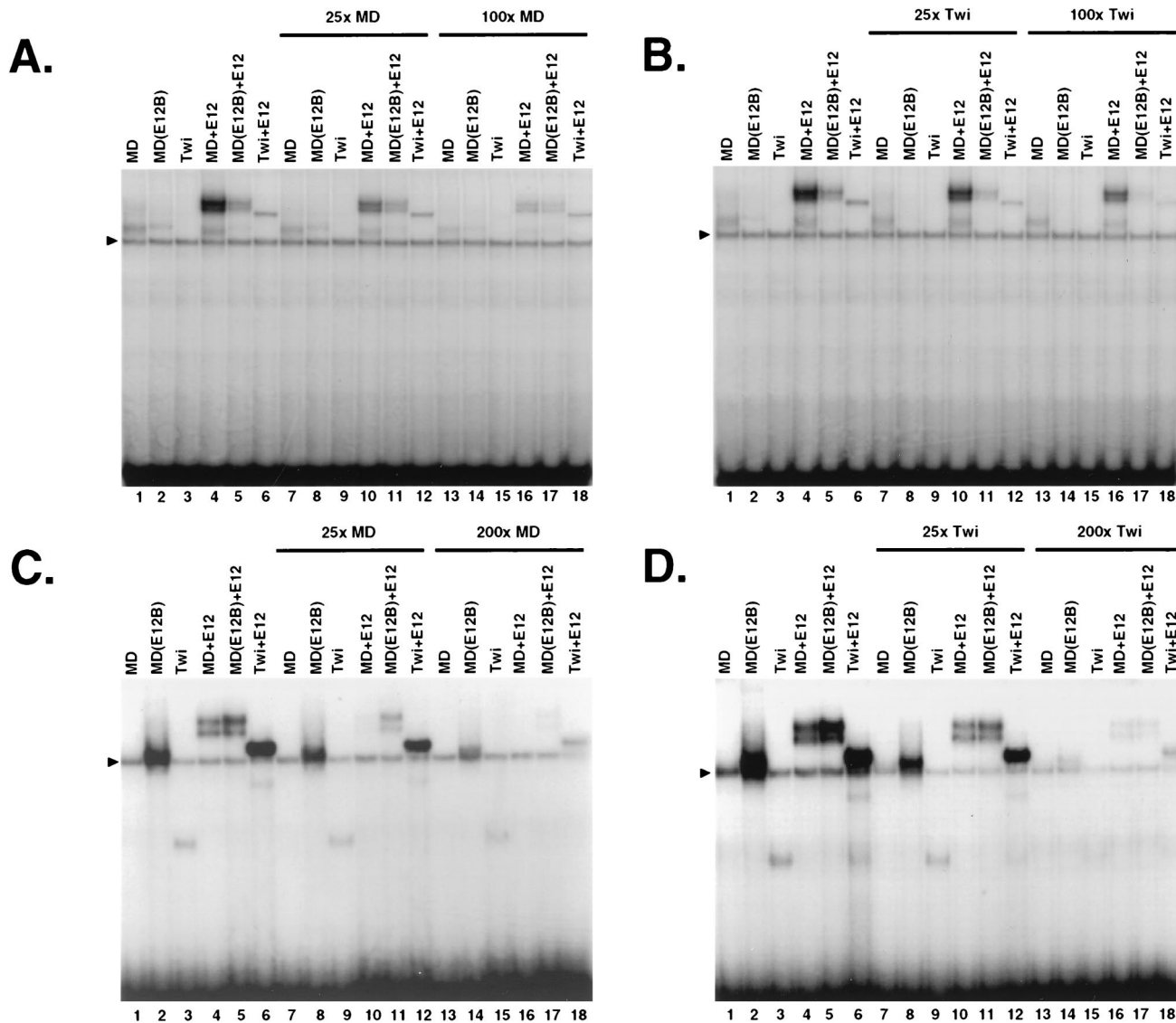


FIG. 6. Binding competition analysis of DNA binding by bHLH heterodimers. (A and B) Binding of the indicated protein complexes to the labeled MyoD site (Fig. 3D) was competed by addition of an unlabeled binding site at ratios indicated above the gel. These experiments were performed and analyzed as for Fig. 4 except that labeled DNA was present at 600 pM, E12 protein present at 8 pM, and all other proteins were present at 19 pM. Twi, Twist. (C and D) Binding of the indicated protein complexes to the labeled Twist site (Fig. 3D) was competed by addition of the indicated unlabeled sites. These experiments were performed described for panel A and B except that labeled DNA was present at 1.1 nM, and they were analyzed by autoradiography. Note that the gel shown in panel C was exposed longer than that shown in panel D, as indicated by comparison of lanes 1 to 6. A background species is indicated by a triangle.

MD(E12B)-E12(MDB), Twist, and Twist-E12 complexes (Fig. 5C, lanes 6, 8, 9, 15, 17, 18, 24, 26, and 27).

Binding site competition and protein titration assays also supported the *in vitro* selection data. The MyoD site competed more effectively than the Twist site for binding by either MyoD homodimers or MyoD-E12 heterodimers (Fig. 6A and B, lanes 1, 4, 7, 10, 13, and 16). In contrast, the Twist site competed more effectively for binding by MD(E12B), MD(E12B)-E12, Twist, and Twist-E12 complexes, although these latter complexes appeared to bind with less specificity than did MyoD-E12 complexes (Fig. 6C and D, lanes 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, and 18). However, the distinct binding specificities of MyoD-E12 and Twist-E12 heterodimers were apparent in a protein titration assay in which the amount of MyoD or Twist protein was varied under conditions of low DNA concentration

(Fig. 7A and B, lanes 1 to 6 and 13 to 18) that more closely represent differences in binding affinity (13). Also in agreement with results described above (Fig. 5C, lanes 14 and 23), heterodimers of MD(E12B) plus E12 bind to the MCK-R site with decreased specificity and with slightly lower affinity than MyoD-E12 complexes (Fig. 7A and B, lanes 7 to 12).

To investigate the role of the BR-HLH junction region in BR positioning, we examined the DNA binding preferences of the MD(E12BJ) and E12(MDBJ) mutants, each of which contains both the BR and junction of the other partner (Fig. 2). In contrast to MD(E12B)-E12(MDB) heterodimers (Fig. 5A and B; Fig. 5C, lanes 6, 15, and 24), MD(E12BJ)-E12(MDBJ) heterodimers (Fig. 2A) bound to the MyoD, Twist, and MCK-R sites with relative preferences that are comparable to those of MD-E12 heterodimers (Fig. 5C, lanes 3, 7, 12, 16, 21,

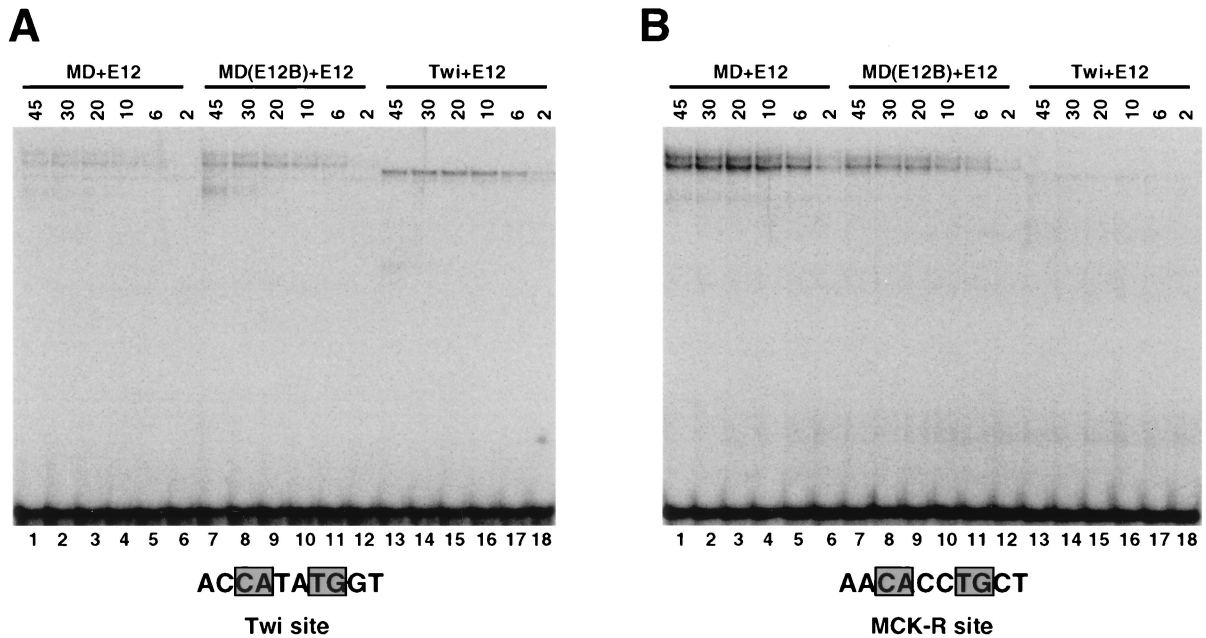


FIG. 7. Protein titration of DNA binding by bHLH heterodimers. (A) Binding to the Twist (Twi) site, analyzed by EMSA and phosphorimaging. In each experiment, E12 was present at 8 pM and DNA that had been labeled to the same specific activity was present at 5 pM. The indicated partner proteins were present at the concentrations (picomolar) shown above the gel. (B) Binding to the MCK-R site, analyzed as for panel A.

and 25). Apparently, the Twist-like sequence preference resulting from simultaneous mispairing of both the MyoD and E12 BRs (Fig. 5A and B) can be corrected by matching each of these BRs with the corresponding junction region. Similarly, and in contrast to MD(E12B) homodimers, MD(E12BJ) homodimers bind to the MyoD, Twist, and MCK-R sites with preferences that are similar to those of E2A proteins (Fig. 8B and C, lane 20, and data not shown). These findings indicate that the BR-HLH junction can be critical for establishing the sequence specificity of an E2A BR, presumably because it influences how the BR is positioned on the DNA.

Contributions of the BR and junction to binding affinity and specificity. It has been shown previously that introduction of A₅, T₆, and either the junction region or K₁₅ of MyoD confers upon E12 the capacity to induce myogenesis (Fig. 2) (18). In the MyoD-DNA complex, A₅ and T₆ are not positioned to allow direct protein-protein contact (Fig. 1) (35), but we have shown that they are critical for the DNA sequence preferences of MyoD, apparently because they affect the conformation of the BR-DNA complex. We have also determined that the junction region can influence how the E2A BR binds DNA. These observations suggest the possibility that the capacity for myogenesis derives entirely from the conformation of the DNA-bound MyoD BR, a model which would predict that the sequence preference of each of these bHLH proteins is established by amino acids at BR positions 5, 6, and 15. We have investigated this model by determining how individual substitutions at these positions, which have been shown to be critical *in vivo*, influence the DNA binding preferences of MyoD.

To address the importance of the MyoD junction region for DNA binding, we altered MyoD positions 14 and 15 (Fig. 8A) and left position 13 intact because it is not required for the MyoD sequence preference in the MD(E12B-AT) mutant (Fig. 2 and 3C). Substitution of alanine for S₁₄, which does not interact with DNA (35), increased binding affinity [MD(AK) (Fig. 8A; Fig. 8B and C, lanes 4 and 5)], perhaps by stabilizing

the BR helix. The preference of MD(AK) for the MyoD site was not substantially altered by replacement of position 15 with alanine [MD(AA)] or with either glutamic acid [MD(AD)] or serine [MD(AS) and MD(OS)], which correspond to residues from E12 or Twist, respectively (Fig. 8; Fig. 8B and C, lanes 5 to 9). The relative preferences of these mutants for the MyoD site are comparable to the binding preferences of other proteins that were confirmed by binding competition analysis (Fig. 4 and 6). Apparently, appropriately specific DNA binding by MyoD homodimers is not impaired by a variety of BR-HLH junction substitutions, including nonconservative mutations of K₁₅. This flexibility contrasts with the importance of the junction region for positioning the E12 BR and with the requirement for K₁₅ for myogenesis.

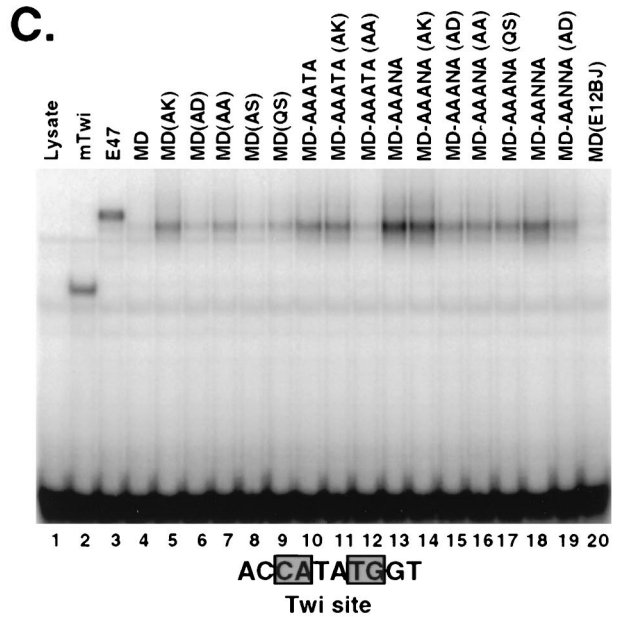
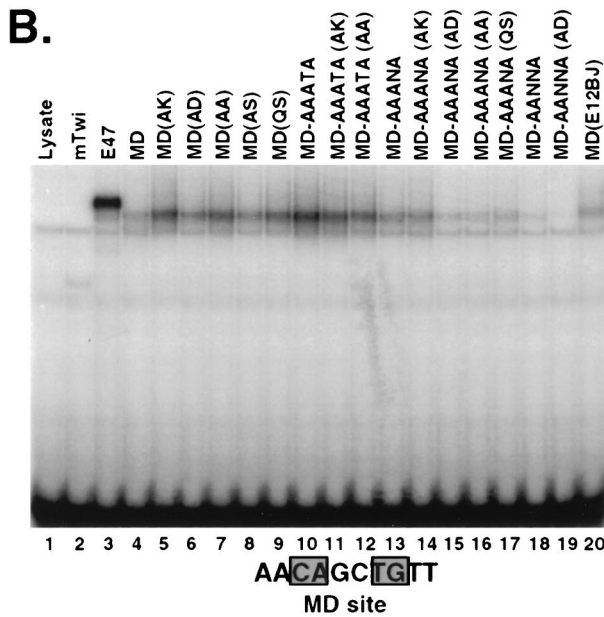
To investigate the role of BR positions 5 and 6 in a neutral context, we first substituted alanine for two nonconserved BR residues (MD-AAATA [Fig. 8A]) that are not predicted to be required for DNA binding (22, 35). This substitution proportionally increased binding to both sites in the context of MyoD (MD-AAATA [Fig. 8B and C, lanes 10]) and enhanced specificity for the MyoD site in the context of MD(AA) (Fig. 8A; Fig. 8B and C, lanes 12). Replacement of T₆ with asparagine conferred a preference for the Twist site (MD-AAANA [Fig. 8A; Fig. 8B and C, lanes 10 and 13]), a finding that parallels the preferences of MD(E12B-AT) and MD(E12B-A) (Fig. 3B and C). This effect was not diminished by various BR-HLH junction mutations or enhanced by presence of Twist junction residues (Fig. 8B and C, lanes 13 to 17), indicating that N₆ is the most important of these residues for the Twist sequence preference. To test whether E2A amino acids that correspond to the three myogenic residues could specify an E2A-like DNA binding preference, we introduced an asparagine at BR position 7 into MD-AAANA and MD-AAANA(AD), the latter of which contains the D₁₅ residue characteristic of E2A proteins (Fig. 8A). In contrast to MD(E12BJ), these mutants strongly preferred the Twist site to the MyoD or MCK-R sites (Fig. 8B

A

MyoD	R	R	K	A	A	T	M	R	E	R	R	R	L	S	K	V
E12	R	R	V	A	N	N	A	R	E	R	L	R	V	R	D	I
Twist	Q	R	V	M	A	N	V	R	E	R	Q	R	T	Q	S	L
MD-AAATA	R	R	<u>A</u>	A	A	T	<u>A</u>	R	E	R	R	R	L	S	K	V
MD-AAANA	R	R	<u>A</u>	A	A	<u>N</u>	<u>A</u>	R	E	R	R	R	L	S	K	V
MD-AANNA	R	R	<u>A</u>	<u>A</u>	<u>N</u>	<u>N</u>	<u>A</u>	R	E	R	R	R	L	S	K	V
MD-(AK)	R	R	K	A	A	T	M	R	E	R	R	R	L	<u>A</u>	K	V
MD-(AA)	R	R	K	A	A	T	M	R	E	R	R	R	L	<u>A</u>	<u>A</u>	V
MD-(AD)	R	R	K	A	A	T	M	R	E	R	R	R	L	<u>A</u>	<u>D</u>	V
MD-(AS)	R	R	K	A	A	T	M	R	E	R	R	R	L	<u>A</u>	<u>S</u>	V
MD-(QS)	R	R	K	A	A	T	M	R	E	R	R	R	L	<u>Q</u>	<u>S</u>	V
MD-AAATA (AK)	R	R	<u>A</u>	A	A	T	<u>A</u>	R	E	R	R	R	L	<u>A</u>	K	V
MD-AAATA (AA)	R	R	<u>A</u>	A	A	T	<u>A</u>	R	E	R	R	R	L	<u>A</u>	<u>A</u>	V
MD-AAANA (AK)	R	R	<u>A</u>	A	A	<u>N</u>	<u>A</u>	R	E	R	R	R	L	<u>A</u>	K	V
MD-AAANA (AA)	R	R	<u>A</u>	A	A	<u>N</u>	<u>A</u>	R	E	R	R	R	L	<u>A</u>	<u>A</u>	V
MD-AAANA (AD)	R	R	<u>A</u>	A	A	<u>N</u>	<u>A</u>	R	E	R	R	R	L	<u>A</u>	<u>D</u>	V
MD-AAANA (QS)	R	R	<u>A</u>	A	A	<u>N</u>	<u>A</u>	R	E	R	R	R	L	<u>Q</u>	<u>S</u>	V
MD-AANNA (AD)	R	R	<u>A</u>	<u>A</u>	<u>N</u>	<u>N</u>	<u>A</u>	R	E	R	R	R	L	<u>A</u>	<u>D</u>	V

Basic Region Junction

FIG. 8. Effects of bHLH BR and BR-HLH junction residues on MyoD binding preferences. (A) Mutagenesis analysis of the MyoD BR and junction. MyoD BR mutant sequences are compared with the MyoD, E12, and Twist BR sequences (Fig. 2). Conserved bHLH residues are shaded, and residues that are altered within full-length MyoD are underlined. (B) Binding of MyoD mutants described in panel A to the MyoD preferred site. These mutants are compared with the indicated wild-type proteins, and binding is assayed as for Fig. 3D except that each protein is present at 40 pM and DNA labeled to the same specific activity is present at 400 pM. E47 is an alternatively spliced E2A protein that binds DNA well as a homodimer (40). Twi, Twist. (C) Binding of MyoD mutants to the Twist preferred site, assayed as for panel B.



and C, lanes 18 to 20, and data not shown), indicating that establishment of an E2A homodimer sequence preference requires additional E2A BR or junction residues and that the conformational mechanisms that dictate this asymmetric sequence preference might be complex.

In the examples that we have analyzed, MyoD mutants that lack myogenic activity bind preferentially to the Twist site (Fig. 2 and 3C), raising the question of whether changes in DNA binding preferences accompany conversion of E12 to a myo-

genic protein through introduction of MyoD BR and junction residues. E12 homodimers do not bind DNA as well as the E2A protein E47 (Fig. 9, lanes 1, 2, 8, 9, 15, and 16), which cannot induce myogenesis (18). Introduction of the MyoD BR into E12 is not sufficient for myogenesis [E12(MDB) (Fig. 2)] but sharply increased binding of E12 to all three sites and was associated with a modest preference for the MyoD site (Fig. 9, lanes 3, 10, and 17). The E12(MDBJ) mutant, which can induce myogenesis (Fig. 2), bound to each of the three sites at a

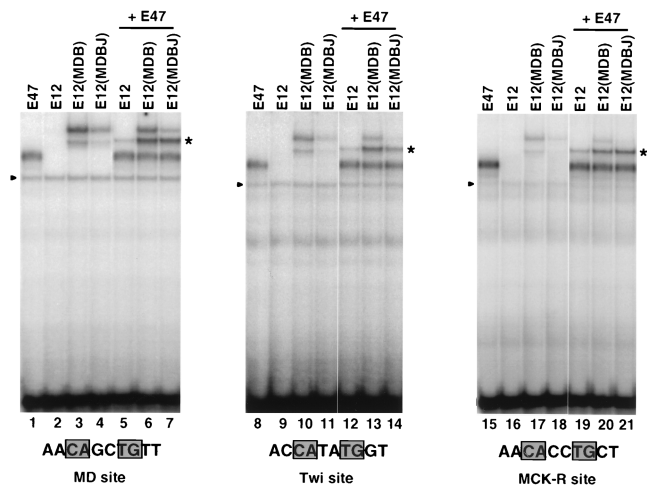


FIG. 9. DNA binding by E12 mutants. DNA binding by the indicated protein complexes is assayed as for Fig. 5C except that all E12 derivatives are present at 8 pM and E47 is present at 19 pM. A protein-DNA complex of intermediate mobility that corresponds to E47-E12 heterodimers is indicated by an asterisk, and a background species is indicated by a closed triangle.

lower level than E12(MDB) and did not have a markedly increased preference for either the MyoD or MCK-R sites (Fig. 9, lanes 4, 11, and 18). Heterodimerization with E47 increased the relative levels with which E12(MDBJ) bound to the MyoD and MCK-R sites (Fig. 9, lanes 6, 7, 13, 14, 20, and 21) but also did not identify DNA binding effects that appear to be sufficient to account for the different functional properties of E12(MDB) and E12(MDBJ). These findings further support the idea that the MyoD junction region is not critical for DNA binding (Fig. 8B and C, lanes 4 to 9) and instead is important for myogenesis because it is involved in other interactions (18).

DISCUSSION

bHLH protein DNA binding specificity deriving from effects on BR-DNA conformation. The myogenic MyoD BR residues A_5 and T_6 are essential for myogenesis but not for binding of MyoD-E2A heterodimers to a muscle-specific site in vitro or in vivo (18, 57). However, we have determined that these residues are required for MyoD to bind DNA with its characteristic specificity for particular CAN NTG sites. Substitution of asparagine for T_6 , and especially for both A_5 and T_6 , results in MyoD binding preferentially to a Twist site (Fig. 8B and C, lanes 10, 13, and 18). The Twist-like MD(E12B) sequence preference is affected partially by substitution of A_5 for the corresponding asparagine [MD(E12B-A) (Fig. 3C)] but is reconfigured by introduction of both A_5 and T_6 so that it is indistinguishable from that of wild-type MyoD [MD(E12B-AT), Fig. 3C]. The data indicate that MyoD residues A_5 and T_6 are each critical for its DNA binding sequence preferences and that the N_6 residue, which is common to the Twist and MD(E12B-A) BRs (Fig. 2), is important for the Twist-like preference. Mutations of these individual BR residues alter sequence preferences across each half-site (Fig. 3C), raising the question of how they might have such a global effect on how the BR helices and the DNA interact preferentially with each other.

A structure of MyoD obtained by X-ray crystallography suggests how A_5 and T_6 might influence binding sequence specificity. When bound to its preferred recognition site, MyoD

does not directly contact base pairs that it specifies in the center of and flanking the CAN NTG consensus (35). However, A_5 and T_6 allow the MyoD BR helix to pack more tightly into the major groove than do the corresponding N_5 and N_6 residues of E2A proteins, in part because of their smaller sizes (Fig. 1 and 2) (35). As a result, the MyoD BR residues T_6 and R_2 directly contact CAN NTG bases at ± 2 and ± 3 respectively, and R_1 binds a backbone phosphate at ± 6 (Fig. 1) (35). In contrast, in E47 R_2 swings out of the major groove and contacts the backbone, and the residue at position 1 does not interact directly with the DNA (12, 19). Supporting the idea that A_5 and T_6 influence the conformation of the DNA-bound BR, substitution of asparagine for A_5 in MyoD increases its sensitivity to protease digestion (29). Our findings suggest that protein-DNA interactions that depend specifically on the MyoD A_5 and T_6 residues may directly influence how the BR helix interacts preferentially with the DNA and thereby indirectly specify its characteristic sequence preferences at positions within and flanking the CAN NTG consensus.

Such indirect conformational effects also appear to be critical for the E2A and Twist sequence preferences. When E47 homodimers bind DNA, a single subunit contacts a base in the center of the site through R_{10} (Fig. 2). This interaction could be important for the asymmetric E2A homodimer sequence preference (19). However, the Twist-like sequence preference that is characteristic of Twist-E2A heterodimers and MD(E12B) homodimers is different across each 5-bp half-site and symmetric (Fig. 3C and 5B), suggesting that it is likely to be established indirectly, through an intermolecular effect that involves a distinct positioning of the BR helix. Introduction of the E12 BR-HLH junction region into MD(E12B) corrects its binding preference so it is like that of E2A homodimers [MD(E12BJ) (Fig. 5C, lanes 7, 16, and 25; Fig. 8B and C, lanes 20)], implicating the BR-HLH junction in this effect. Presumably, the E2A junction acts in concert with the asparagines at BR positions 5 and 6 (Fig. 2), although the Twist-like preference of the MD-AA₅NA(AD) mutant (Fig. 8B and C, lane 19, and data not shown) suggests that the E2A junction residue D_{15} is not sufficient. The finding that E2A proteins can be targeted to different DNA sequences by different dimer partners may have important implications for their in vivo functions.

In contrast, the BR-HLH junction region does not have a strong influence on the MyoD DNA binding preference. Various MyoD junction mutations do not substantially diminish its preference for a MyoD site (Fig. 8B and C, lanes 5 to 9). In addition, the similar sequence preferences of E12(MDB) and E12(MDBJ) homodimers (Fig. 9, lanes 3, 4, 10, 11, 17, and 18) contrast sharply with the different specificities of MD(E12B) and MD(E12BJ) (Fig. 3D, lanes 2 and 6; Fig. 8B and C, lanes 20). This apparent difference between MyoD and E2A proteins might derive from the distinct arrangement of the BR helix on the DNA that results from presence of MyoD residues A_5 and T_6 .

It is striking that as a group, these various bHLH mutants and dimer combinations bind DNA with a limited number of discrete sequence preferences (Figs. 3C and 5B). Presumably, each of these preferences reflects a preferred conformational state that is dictated by how each BR helix and the corresponding DNA sequence conform to each other in an induced fit (49). This mechanism for recognizing particular CAN NTG sites appears to be different from the direct recognition of central bases that is characteristic of bHLH proteins that contain R_{13} and bind to CACGTG or CATGTG sites (20, 21, 48). Consistent with this idea, BR residues 5 and 6 do not appear to be important for the function of the R_{13} -containing bHLH

protein c-Myc (10). In E2A and its tissue-specific dimerization partners, a more flexible conformation-based mechanism might have evolved to increase adaptability in both sequence recognition and function, so that different combinations of these proteins can result in distinct protein-DNA conformations that correspond to particular DNA sequence preferences. Such a model may be particularly plausible for bHLH proteins, because folding of the BR into an α helix is driven by its interaction with the DNA (2).

BR-DNA conformation, DNA binding specificity, and myogenesis. The observation that the MyoD junction and K_{15} are not required for an appropriate DNA binding specificity (Fig. 8B and C, lanes 6 to 9; Fig. 9) supports the model that K_{15} is involved in other essential interactions (18). However, our experiments also pose the question of how the functional importance of A_5 and T_6 might be related to their effects on DNA recognition. Of the MyoD BR mutants that we have analyzed, those that do not induce myogenesis bind to DNA as homodimers with a Twist-like preference [MD(E12B) and MD(E12B-A) (Fig. 2 and 3C)]. Heterodimers of MD(E12B) with E12 prefer a heterodimer site (Fig. 5B), but with markedly diminished specificity compared to MyoD-E12 dimers (Fig. 5C, lanes 3, 5, 12, 14, 21, and 23; Fig. 6; Fig. 7A and B, lanes 1 to 12). This finding suggests that at least in part, A_5 and T_6 may be significant for myogenesis because they restrict the DNA binding specificity of MyoD and other myogenic bHLH proteins, so that they are less likely to bind inappropriate sites. However, other observations support a role for the A_5 and T_6 residues in protein-protein interactions. They have been implicated in binding to other proteins when they are not bound to DNA (26, 38), and evidence indicates that they are required for activation domain exposure (5, 29, 57) and cooperative DNA binding (3). Finally, unlike MyoD, MD(E12B) can activate transcription of a reporter only in particular cell lines, implicating the BR in protein-protein interactions (57).

In light of evidence that A_5 and T_6 establish the conformation of the DNA-bound BR, it is an attractive model that this effect might influence the function of myogenic bHLH proteins directly, by affecting their interactions with other proteins. Given that relatively subtle alterations of the MyoD BR and junction region can enhance MyoD DNA binding significantly [MD(AK) and MD(AAATA) (Fig. 8B and C, lanes 4, 5, and 10)], it appears likely that cooperative protein-protein interactions with the BR and junction could influence binding affinity. It has been demonstrated recently that MyoD binds cooperatively with other DNA binding proteins to a particular muscle-specific promoter (4). The E box sequences through which MyoD activates transcription in the context of this promoter can differ from those that it binds preferentially *in vitro* (28), suggesting that DNA sequence recognition may be influenced by interactions with cooperating proteins *in vivo*. In addition, interactions with cooperating proteins might be influenced in turn by the specificity of DNA sequence recognition, as suggested by evidence that for MyoD and E proteins, the choice between homo- or heterodimer formation may be dictated by the DNA binding affinities of the individual BRs (36, 59). Our findings are consistent with the idea that deceptively subtle aspects of sequence recognition could be important for the biological activity of MyoD, if they influence functionally critical interactions that might also involve K_{15} or other MyoD regions.

ACKNOWLEDGMENTS

For critically reading the manuscript we thank Robert Davis, Stephen Tapscott, members of the Blackwell laboratory, and Tom Ellenberger, whom we also thank for help with computer graphics. We thank Lau-

ren Snider for advice on bacterial expression and partial purification of MyoD, and we thank Amy Chen and Jonathan Mitchell for contributing to early stages of this project. The *Xenopus* Twist cDNA was provided by Ralph Rupp, and the murine Twist cDNA was donated by Doug Spicer and Andrew Lassar. This work was initiated in the laboratory of Harold Weintraub, to whom T.K.B. is particularly grateful for support, advice, and inspiration.

This work was supported by grant DAMD17-94-J-4063 from the U.S. Army Breast Cancer Program to T.K.B.

REFERENCES

- Aksan, I., and C. R. Goding. 1998. Targeting the microphthalmia basic helix-loop-helix-leucine zipper transcription factor to a subset of E-box elements *in vitro* and *in vivo*. *Mol. Cell. Biol.* **18**:6930–6938.
- Anthony-Cahill, S. J., P. A. Benfield, R. Fairman, Z. R. Wasserman, S. L. Brenner, W. F. Stafford, C. Altenbach, W. L. Hubbell, and W. F. DeGrado. 1992. Molecular characterization of helix-loop-helix peptides. *Science* **255**: 979–983.
- Bengal, E., O. Flores, P. N. Rangarajan, A. Chen, H. Weintraub, and I. M. Verma. 1994. Positive control mutations in the MyoD basic region fail to show cooperative DNA binding and transcriptional activation *in vitro*. *Proc. Natl. Acad. Sci. USA* **91**:6221–6225.
- Biesiada, E., Y. Hamamori, L. Keddes, and V. Sartorelli. 1999. Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multi-protein transcriptional complex required for activity of the human cardiac alpha-actin promoter. *Mol. Cell. Biol.* **19**:2577–2584.
- Black, B. L., J. D. Molkenin, and E. N. Olson. 1998. Multiple roles for the MyoD basic region in transmission of transcriptional activation signals and interaction with MEF2. *Mol. Cell. Biol.* **18**:69–77.
- Blackwell, T. K. 1995. Selection of protein binding sites from random nucleic acid sequences. *Methods Enzymol.* **254**:604–618.
- Blackwell, T. K., J. Huang, A. Ma, L. Kretzner, F. W. Alt, R. N. Eisenman, and H. Weintraub. 1993. Binding of Myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell. Biol.* **13**:5216–5224.
- Blackwell, T. K., L. Kretzner, E. M. Blackwood, R. N. Eisenman, and H. Weintraub. 1990. Sequence-specific DNA binding by the c-Myc protein. *Science* **250**:1149–1151.
- Blackwell, T. K., and H. Weintraub. 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**:1104–1110.
- Bodis, S., T. Hemesath, and D. E. Fisher. 1997. Highly conserved asparagine in the basic domain of Myc is dispensable for DNA binding, transformation, and apoptosis. *Biochem. Mol. Med.* **60**:102–107.
- Brennan, T. J., T. Chakraborty, and E. N. Olson. 1991. Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. *Proc. Natl. Acad. Sci. USA* **88**:5675–5679.
- Brownlie, P., T. Ceska, M. Lamers, C. Romier, G. Stier, H. Teo, and D. Suck. 1997. The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* **5**:509–520.
- Carey, J. 1991. Gel retardation. *Methods Enzymol.* **208**:103–117.
- Chakraborty, T., T. J. Brennan, L. Li, D. Edmondson, and E. N. Olson. 1991. Inefficient homo-oligomerization contributes to the dependence of myogenin on E2A products for efficient DNA binding. *Mol. Cell. Biol.* **11**:3633–3641.
- Cripps, R. M., B. L. Black, B. Zhao, C. L. Lien, R. A. Schulz, and E. N. Olson. 1998. The myogenic regulatory gene Mef2 is a direct target for transcriptional activation by Twist during *Drosophila* myogenesis. *Genes Dev.* **12**:422–434.
- Dang, C. V., C. Dolde, M. L. Gillison, and G. J. Kato. 1992. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc. Natl. Acad. Sci. USA* **89**:599–602.
- Davis, R. L., P. F. Cheng, A. B. Lassar, and H. Weintraub. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60**:733–746.
- Davis, R. L., and H. Weintraub. 1992. Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science* **256**:1027–1030.
- Ellenberger, T., D. Fass, M. Arnaud, and S. C. Harrison. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* **8**:970–980.
- Ferre-D'Amare, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley. 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363**:38–45.
- Ferre-D'Amare, A. R., P. Pognonec, R. G. Roeder, and S. K. Burley. 1994. Structure and function of the b/HLH/Z domain of USF. *EMBO J.* **13**:180–189.
- Fisher, D. E., L. A. Parent, and P. A. Sharp. 1993. High affinity DNA-binding Myc analogs: recognition by an α helix. *Cell* **72**:467–476.
- Fisher, F., D. H. Crouch, P.-S. Jayaraman, W. Clark, D. A. F. Gillespie, and C. R. Goding. 1993. Transcription activation by Myc and Max: flanking sequences target activation to a subset of CACGTG motifs *in vivo*. *EMBO J.* **12**:5075–5082.

24. **Gould, K. A., and E. H. Bresnick.** 1998. Sequence determinants of DNA binding by the hematopoietic helix-loop-helix transcription factor TAL1: importance of sequences flanking the E-box core. *Gene Expr.* **7**:87–101.
25. **Halazonetis, T. D., and A. N. Kandil.** 1992. Predicted structural similarities of the DNA binding domains of c-Myc and endonuclease Eco RI. *Science* **255**:464–466.
26. **Hamamori, Y., H. Y. Wu, V. Sartorelli, and L. Keddes.** 1997. The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, Twist. *Mol. Cell. Biol.* **17**:6563–6573.
27. **Harfe, B. D., C. S. Branda, M. Krause, M. J. Stern, and A. Fire.** 1998. MyoD and the specification of muscle and non-muscle fates during postembryonic development of the *C. elegans* mesoderm. *Development* **125**:2479–2488.
28. **Huang, J., T. K. Blackwell, L. Keddes, and H. Weintraub.** 1996. Differences between MyoD DNA binding and activation site requirements revealed by a functional random sequence selection. *Mol. Cell. Biol.* **16**:3893–3900.
29. **Huang, J., H. Weintraub, and L. Keddes.** 1998. Intramolecular regulation of MyoD activation domain conformation and function. *Mol. Cell. Biol.* **18**:5478–5484.
30. **Jennings, B. H., D. M. Tyler, and S. J. Bray.** 1999. Target specificities of *Drosophila* enhancer of split basic helix-loop-helix proteins. *Mol. Cell. Biol.* **19**:4600–4610.
31. **Kophengnavong, T., A. S. Carroll, and T. K. Blackwell.** 1999. The SKN-1 amino-terminal arm is a DNA specificity segment. *Mol. Cell. Biol.* **19**:3039–3050.
32. **Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub.** 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* **66**:305–315.
33. **Lee, J. E.** 1997. Basic helix-loop-helix genes in neural development. *Curr. Opin. Neurobiol.* **7**:13–20.
34. **Lemerrier, C., R. Q. To, R. A. Carrasco, and S. F. Konieczny.** 1998. The basic helix-loop-helix transcription factor Mist1 functions as a transcriptional repressor of myoD. *EMBO J.* **17**:1412–1422.
35. **Ma, P. C., M. A. Rould, H. Weintraub, and C. O. Pabo.** 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77**:451–459.
36. **Maleki, S. J., C. A. Royer, and B. K. Hurlburt.** 1997. MyoD-E12 heterodimers and MyoD-MyoD homodimers are equally stable. *Biochemistry* **36**:6762–6767.
37. **Michelson, A.** 1996. A new turn (or two) for Twist. *Science* **272**:1449–1450.
38. **Molkentin, J. D., B. L. Black, J. F. Martin, and E. N. Olson.** 1995. Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* **83**:1125–1136.
39. **Molkentin, J. D., and E. N. Olson.** 1996. Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. *Proc. Natl. Acad. Sci. USA* **93**:9366–9373.
40. **Murre, C., P. S. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**:777–783.
41. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**:537–544.
42. **Neuhold, L., and B. Wold.** 1993. HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell* **74**:1033–1042.
43. **Olson, E. N., and W. H. Klein.** 1994. bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. *Genes Dev.* **8**:1–8.
44. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**:1053–1095.
45. **Parraga, A., L. Bellolell, A. R. Ferre-D'Amare, and S. K. Burley.** 1998. Cocystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* **6**:661–672.
46. **Postigo, A. A., and D. C. Dean.** 1999. ZEB represses transcription through interaction with the corepressor CtBP. *Proc. Natl. Acad. Sci. USA* **96**:6683–6688.
47. **Schwarz, J. J., T. Chakraborty, J. Martin, J. M. Zhou, and E. N. Olson.** 1992. The basic region of myogenin cooperates with two transcription activation domains to induce muscle-specific transcription. *Mol. Cell. Biol.* **12**:266–275.
48. **Shimizu, T. X., A. Toumoto, K. Ihara, M. Shimizu, Y. Kyogoku, N. Ogawa, Y. Oshima, and T. Hakoshima.** 1997. Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. *EMBO J.* **16**:4689–4697.
49. **Spolar, R. S., and M. T. Record, Jr.** 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**:777–784.
50. **Steitz, T. A.** 1990. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* **23**:205–280.
51. **Studier, F. W.** 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219**:37–44.
52. **Szymanski, P., and M. Levine.** 1995. Multiple modes of dorsal-bHLH transcriptional synergy in the *Drosophila* embryo. *EMBO J.* **14**:2229–2238.
53. **Thayer, M. J., and H. Weintraub.** 1993. A cellular factor stimulates the DNA-binding activity of MyoD and E47. *Proc. Natl. Acad. Sci. USA* **90**:6483–6487.
54. **Van Antwerp, M. E., D. G. Chen, C. Chang, and E. V. Prochownik.** 1992. A point mutation in the MyoD basic domain imparts c-Myc-like properties. *Proc. Natl. Acad. Sci. USA* **89**:9010–9014.
55. **Voronova, A., and D. Baltimore.** 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proc. Natl. Acad. Sci. USA* **87**:4722–4726.
56. **Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benzeza, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar.** 1991. The MyoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251**:761–766.
57. **Weintraub, H., V. J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S. J. Tapscott.** 1991. Muscle-specific transcriptional activation by MyoD. *Genes Dev.* **5**:1377–1386.
58. **Weintraub, H., T. Genetta, and T. Kadesch.** 1994. Tissue-specific gene activation by MyoD: determination of specificity by cis-acting repression elements. *Genes Dev.* **8**:2203–2211.
59. **Wendt, H., R. M. Thomas, and T. Ellenberger.** 1998. DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J. Biol. Chem.* **273**:5735–5743.
60. **Yin, Z., X. L. Xu, and M. Frasch.** 1997. Regulation of the twist target gene tinman by modular cis-regulatory elements during early mesoderm development. *Development* **124**:4971–4982.