DNA Binding Specificity of the Basic−Helix−Loop−Helix Protein MASH-1†

Daniel Meierhans, Chirine el-Ariss, Margrit Neuenschwander, Martin Sieber, Joseph F. Stackhouse, and Rudolf K. Allemann*

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

Received April 24, 1995; Revised Manuscript Received June 26, 1995®

ABSTRACT: Despite the high degree of sequence similarity in their basic−helix−loop−helix (BHLH) domains, MASH-1 and MyoD are involved in different biological processes. In order to define possible differences between the DNA binding specificities of these two proteins, we investigated the DNA binding properties of MASH-1 by circular dichroism spectroscopy and by electrophoretic mobility shift assays (EMSA). Upon binding to DNA, the BHLH domain of MASH-1 underwent a conformational change from a mainly unfolded to a largely α-helical form, and surprisingly, this change was independent of the specific DNA sequence. The same conformational transition could be induced by the addition of 20% 2,2,2-trifluoroethanol. The apparent dissociation constants (K_D) of the complexes of full-length MASH-1 with various oligonucleotides were determined from half-saturation points in EMSAs. MASH-1 bound as a dimer to DNA sequences containing an E-box with high affinity ($K_D = 1.4-4.1 \times 10^{-14}$ M^2). However, the specificity of DNA binding was low. The dissociation constant for the complex between MASH-1 and the highest affinity E-box sequence ($K_D = 1.4 \times 10^{-14}$ M^2) was only a factor of 10 smaller than for completely unrelated DNA sequences ($K_D = \sim 1 \times 10^{-13}$ M^2). The DNA binding specificity of MASH-1 was not significantly increased by the formation of an heterodimer with the ubiquitous E12 protein. MASH-1 and MyoD displayed similar binding site preferences, suggesting that their different target gene specificities cannot be explained solely by differential DNA binding. An explanation for these findings is provided on the basis of the known crystal structure of the BHLH domain of MyoD.

MASH-1 (Figure 1A) belongs to a family of transcription factors, many members of which are involved in processes regulating cellular commitment and differentiation (Campuzano & Modolell, 1992; Johnson et al., 1990; Lo et al., 1991; Weintraub, 1993). The expression of MASH-1 is required for the early development of the olfactory and the autonomic neurons (Lo et al., 1991), while the expression of MyoD, myogenin, Myf-5, and MRF-4 can activate myogenesis in myoblasts as well as in other cell types (Weintraub et al., 1989; Olson, 1990). The myogenic proteins play a central role in determining muscle cell identity (Hasty et al., 1993; Nabeshima et al., 1993; Rudnicki et al., 1993) and have sometimes been called master regulators of muscle differentiation (Paterson et al., 1991; Olson, 1990). The functional activity of these myogenic proteins depends on heterodimerization with E12 or E47, the products of the ubiquitously expressed E2A gene (Lassar et al., 1991). It has been suggested that MASH-1 also heterodimerizes with the E12/47 proteins (Johnson et al., 1992).

All of these myogenic and neurogenic transcription factors contain a region of high sequence similarity called the basic−helix−loop−helix (BHLH)[1] motif (Figure 1B) (Murre et al., 1989). Outside this BHLH region, the proteins have little sequence similarity. An X-ray structure of the BHLH region of MyoD shows that the HLH domain mediates interactions between the monomers and that the DNA is contacted mainly through residues of the basic region (Ma et al., 1994). The activity of the BHLH proteins depends upon the presence of a DNA sequence containing the core motif CANNTG (E-box) (Lassar et al., 1989). Such an E-box occurs on the average every 256 base pairs in the genome, and the strict biological specificity of the neurogenic and myogenic factors presents therefore a major paradox. Results obtained in SAAB and CASTing experiments have been interpreted to suggest that the specificity might be achieved by subtle control of the interactions of the BHLH domains with the central two base pairs of the E-box and with the nucleotides flanking the E-box, whereby small changes in the structure of the BHLH domain could induce different phenotypic changes (Blackwell & Weintraub, 1990; Wright et al., 1991; Weintraub et al., 1991).

In order to address the molecular basis of the different functions of the myogenic and the neurogenic regulators, we have produced, in *E. coli*, MASH-1, E12, and the BHLH domain of MASH-1 and determined the DNA binding properties of these transcriptional regulators. We show that dimeric MASH-1 binds to DNA containing an E-box with high affinity. Measurements of the apparent dissociation constants of the complexes of the MASH-1 homodimer and of the MASH-1/E12 heterodimer with several different E-box-containing DNA sequences revealed a sequence preference of MASH-1 similar to that of the myogenic proteins. The dissociation constants for the complexes of MASH-1 and MASH-1/E12 with DNA sequences not containing E-boxes are only 1 order of magnitude higher than for the complexes with E-box-containing DNA sequences, indicating that MASH-1 produced in bacteria

† This work was supported by an ETH grant (R.K.A.).

* Author to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* August 15, 1995.

[1] Abbreviations: BHLH, basic−helix−loop−helix domain; BSA, bovine serum albumin; CASTing, cyclic amplification and selection of targets; CD, circular dichroism; CHAPS, 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate; EMSA, electrophoretic mobility shift assay; HPLC, high-performance liquid chromatography; MCK, muscle creatine kinase; TFE, 2,2,2-trifluoroethanol; SAAB, selected and amplified binding site "imprints".

0006-2960/95/0434-11026$09.00/0

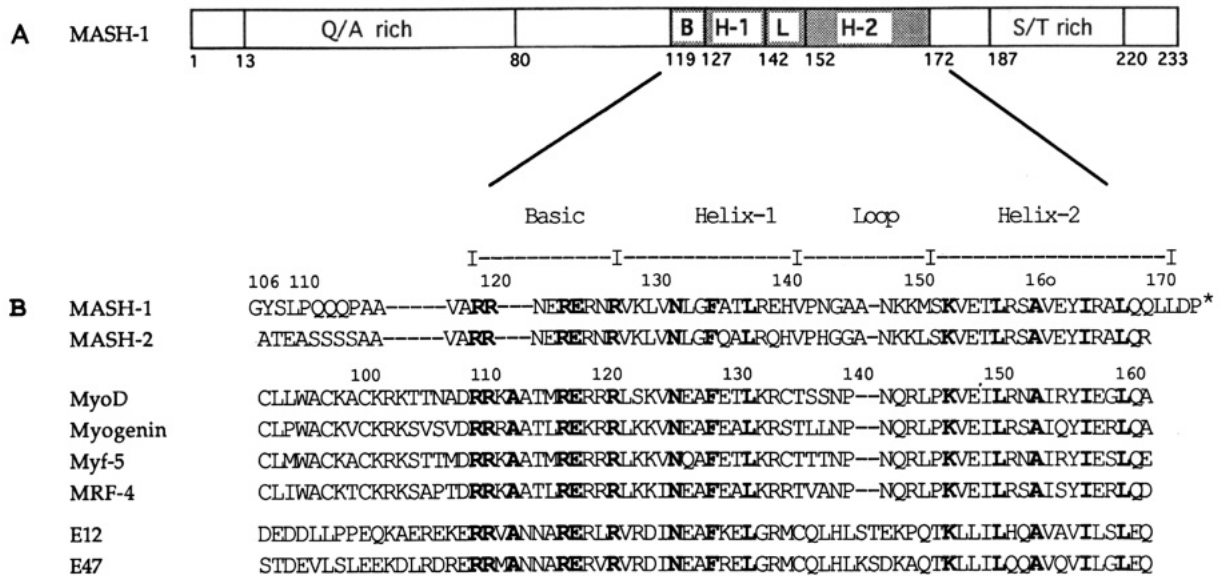

FIGURE 1: (A) Cartoon of the MASH-1 protein. The numbering indicated and used throughout the text corresponds to the full-length sequence of MASH-1. (B) Sequence similarities of the BHLH domains of MASH-1 and of other BHLH proteins. The alignment is grouped into the neurogenic, the myogenic, and the E12/47 subfamilies. Amino acids are given in the one-letter code, and conserved residues are in boldface. The proline residue at the carboxy terminus of the MASH-1 BHLH sequence is due to the cloning procedure and is not part of the MASH-1 sequence. For MyoD the numbering of the full-length MyoD protein was used, and this is indicated above the BHLH sequence of MyoD. The sequences for MASH-1 and -2 (Johnson et al., 1990) and for E12 and E47 (German et al., 1991) are from rat; for MyoD (Davis et al., 1987), myogenin (Edmondson & Olson, 1989), Myf-5 (Buonanno et al., 1992), and MRF-4 (Miner & Wold, 1990), the murine sequences are indicated. Abbreviations: B, basic region; H-1 and H-2, helix 1 and 2, respectively; L, loop region; Q/A and S/T rich, protein stretches rich in Gln and Ala and in Ser and Thr, respectively.

displays only low intrinsic sequence specificity, a property also displayed by the BHLH domain of MASH-1 in the presence of TFE. These results show that unmodified MASH-1 alone cannot be responsible for the specificity needed to distinguish the E-box of a nerve-specific enhancer from an E-box in a DNA region controlling the expression of a muscle specific gene and that the specificity of gene expression must be achieved either through posttranslational modifications of MASH-1, which enhance the specificity of DNA binding, or through specific interplay with other components of the transcriptional machinery, thereby increasing the specificity of gene activation by cooperative interaction.

MATERIALS AND METHODS

Expression of MASH-1, BHLH of MASH-1, and E12. All recombinant proteins were expressed in BL21(DE3) cells containing the pLysS plasmid (Studier, 1991) from the T7 promoter in the plasmid pJGetita, a derivative of pET3a (Studier & Moffat, 1986). The MASH-1 and the E12 cDNAs were from rat (Johnson et al., 1990; Lo et al., 1991). To construct cDNA inserts for the expression plasmids, PCR-mediated site-directed mutagenesis of the cDNAs was used. For the MASH-1 construct, an *Nde*I site was created at the position of the initiator ATG in the MASH-1 cDNA and a *Bgl*II site just 3′ of the MASH-1 termination codon. The PCR-amplified DNA fragment was inserted between the *Nde*I and *Bam*HI sites of pJGetita. The resulting sequence contains the entire MASH-1 coding region (Figure 1A). A fragment of the MASH-1 cDNA (Figure 1B), coding for the BHLH domain from G(106) to D(172), was amplified by PCR in the same manner with an *Nde*I site being created at the position of the initial ATG codon. The PCR primer for the 3′ end, and thus also the final construct, contained an additional codon for a proline at the last position. The same approach was used for the construction of the ATG start codon in the E12 cDNA for the E12 expression plasmid. For the 3′ terminus, a natural *Bam*HI site in the coding region of E12 was used. The stop codon for this construction was in pJGetita, just after the polylinker. The corresponding E12 construct codes for a protein which lacks the last C-terminal 21 amino acids. BL21(DE3) cells containing these plasmids were grown at 37 °C on YT medium with 100 mg/L ampicillin and 30 mg/L chloramphenicol until the OD_{550} reached 0.4. Then IPTG was added to a final concentration of 0.5 mM. Cells were harvested 3 h after induction and frozen at −20 °C.

Protein Purification. The cells were lysed by thawing in the presence of 3−4 mL of water/g of wet cells and 1 mM phenylmethanesulfonyl fluoride. The cells were resuspended by vortexing, and then 2 volumes of lysis buffer (100 mM ammonium acetate, pH 6.7, 100 mM sodium chloride, and 100 mM 2-mercaptoethanol) was added. The suspension was sonicated for 10 min at room temperature and then centrifuged at 15 000 rpm for 15 min. The pellet was resuspended in lysis buffer (5−10 mL/g of cells) and recentrifuged. This process was repeated once; then the pellet was washed twice with 5 mM sodium acetate (5 mL/g of cells) (pH 5.0) and centrifuged.

For the purification of MASH-1, the pellet was dissolved in urea buffer (5 mM sodium acetate, pH 5.0, 100 mM 2-mercaptoethanol, and 8 M urea) and the solution applied to a column containing 30 mL of Bio-Gel CM A ion-exchange resin (Bio-Rad) preequilibrated with urea buffer. The column was washed extensively with urea buffer and the protein eluted with 2−3 column volumes of urea buffer containing 100 mM sodium chloride. The combined MASH-1 fractions were dialyzed overnight against 20 volumes of 100 mM sodium acetate (pH 5.0), 100 mM sodium chloride, 100

mM 2-mercaptoethanol, and 20% glycerol. This resulted in a white precipitate, with essentially no MASH-1 remaining in the supernatant. The precipitate was collected by centrifugation, redissolved in water, and dialyzed against 10 mM sodium acetate (pH 5.0) and 1 mM DTT. The protein was homogeneous by SDS gel electrophoresis and >95% pure as judged by HPLC analysis on carboxymethyl (Alltech) or sulfonate (Pharmacia) ion-exchange columns. The protein was further purified by preparative HPLC on a Resource-S sulfonate ion-exchange column (Pharmacia) using a linear gradient from 4 M urea in 10 mM sodium acetate (pH 5.0) to 1 M NaCl, 4 M urea, and 10 mM sodium acetate (pH 5.0). The collected fractions were pooled and concentrated by ultrafiltration, and the buffer was exchanged to 10 mM sodium acetate (pH 5.0) and 5 mM DTT. The yields for the preparation were up to 18 mg of purified protein/L of culture. The protein showed a single band by SDS gel electrophoresis and ion-exchange and reverse-phase HPLC. Laser desorption time of flight mass spectrometry showed sample homogeneity and correct molecular mass. The sequence of the first 25 amino acids was determined by automated Edman degradation and agreed with the expected sequence for MASH-1. HPLC analysis by gel filtration (Alltech 4.6 × 250 mm Macrosphere GPC 100, 7 μm) showed the material to be an aggregate at high concentrations. This aggregate was disrupted in the presence of DTT and either 15% acetonitrile or 4% CHAPS. Protein concentrations were determined by measuring the UV absorptions at 215 and 220 nm (Wetlaufer, 1962).

The purification protocol for the BHLH domain of MASH-1 was as above until the dialysis. However, the BHLH domain proved to be soluble in the dialysis buffer and was therefore concentrated by ultrafiltration using an Amicon YM-3 filter, and the buffer was exchanged for 5 mM sodium acetate (pH 5.0). The material eluting from the Bio-Gel CM A column consisted of a mixture of three peptides. Laser desorption time of flight mass spectrometry of the mixture showed masses of 7700.3, 7958.6, and 8930.0. The bands could be separated and the protein further purified by HPLC using a Resource-S sulfonate ion-exchange column with the same buffer system as above. On the basis of their molecular weights, the bands were assigned to the desired BHLH domain of MASH-1, a peptide with 2 more C-terminal amino acids, WA, and a peptide with 11 more C-terminal amino acids, WAITSITPWGL (these peptides were apparently formed by the misreading of the UGA stop codon for UGG in the construct, thus inserting a tryptophan, followed by extension to the next stop codon and limited proteolysis at the C-terminal ends of the expressed proteins). The fractions containing the desired 69 amino acid peptide were pooled and concentrated, and the buffer was exchanged to 5 mM sodium acetate (pH 5.0) and 5 mM DTT. The peptide was homogeneous by mass spectroscopy, SDS gel electrophoresis, and HPLC on reverse-phase and ion-exchange materials.

E12 was purified in a manner identical to that of BHLH up to concentration and buffer exchange by ultrafiltration, except that an Amicon PM-10 filter was used. The material was not applied to a Bio-Gel CM A column but purified directly by HPLC on a Resource-S sulfonate ion-exchange column. The buffers were A, 10 mM sodium acetate (pH 5.0), 50 mM 2-mercaptoethanol, and 8 M urea, and B, 10 mM sodium acetate (pH 5.0), 50 mM 2-mercaptoethanol, 8 M urea, and 1 M NaCl. The proteins were eluted using a gradient from 0% to 3% buffer B over 3 min, followed by linear gradients from 3% to 30% buffer B over 30 min and 30−50% buffer B over 10 min. The collected fractions were pooled and concentrated, and the buffer was exchanged to 5 mM sodium acetate (pH 5.0) and 5 mM DTT. The protein was homogeneous by mass spectroscopy, SDS gel electrophoresis, and HPLC on reverse-phase and ion-exchange materials.

Oligonucleotides. Oligonucleotides were purchased from Microsynth, desalted on Sephadex, and precipitated with ethanol. Sequences of these DNA samples are shown in Figure 2. Single-stranded oligonucleotides were labeled with [^{32}P]ATP (Amersham) in the presence of T4 polynucleotide kinase (New England Biolabs) and complementary strands annealed by heat denaturation followed by slow cooling to room temperature.

Immunological Reagents. Polyclonal antibodies were generated in female New Zealand White rabbits according to standard procedures. Anti-MASH-1 antibodies were further purified by affinity chromatography using purified MASH-1 as an affinity ligand linked to CNBr-activated Sepharose (Pharmacia).

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. Serial dilutions of the stock solutions (0.1 ng to 50 μg) were incubated in 50 mM Tris (pH 7.9), 6 mM $MgCl_2$, 40 mM ammonium sulfate, 0.2 mM EDTA, 1 mM DTT, 0−4% CHAPS, and 0.5−10% glycerol in the presence of competitor DNA (if used) for 15 min at ambient temperature prior to mixing with 50 pg to 4 ng of the labeled oligonucleotide. Binding reactions were allowed to proceed for a minimum of 15 min and a maximum of 120 min at 30 °C. The results were independent of the length of the incubation time. When used, polyclonal antibodies or antisera were added after 10 min of the binding reaction, and the reaction was continued for another 5 min. The samples were applied to a 4% polyacrylamide gel in 0.9 × TAE (pH 7.9), dried, and exposed to Kodak XOMAT-S film at −70 °C. Quantitative data were obtained with a phosphorimager 400 (Molecular Dynamics) using the Image Quant software.

CD Spectroscopy. Spectra were measured on a Jasco J600 circular dichroism spectrophotometer at room temperature. The concentration of BHLH was 1 μM in 5 mM Tris (pH 7.0) and 0−50% TFE. For CD measurements of E12 and of MASH-1, 0.25 mM DTT was also present. For CDs of full-length MASH-1, the pH was adjusted to 6.5 because of the low solubility.

RESULTS

Determination of the Specificity of DNA Binding of MASH-1. The DNA binding properties of MASH-1 were examined by electrophoretic mobility shift assays. As a DNA probe, we chose an oligonucleotide comprising 17 base pairs of the IgH enhancer like element of the muscle-specific creatine kinase enhancer (Buskin & Hauschka, 1989) with the central E-box sequence CAGGTG (Figure 2). This enhancer of a gene involved in myogenesis is not the natural target of the neurogenic factor MASH-1. However, the interactions of the members of the myogenic subfamily of BHLH proteins

	1 2 3 4 5 6
MCK-S(5'G)	5'-CAGGCA **G C A G G T G** TTGG-3'
	3'-GTCCGT **C G T C C A C** AACC-5'
	6'5'4'3'2'1'
E-BoxCA	5'-CAGGCA **G C A c a T G** TTGG-3'
E-BoxTC	5'-CAGGCA **G C A t c T G** TTGG-3'
E-BoxAC	5'-CAGGCA **G C A a c T G** TTGG-3'
E-BoxGC	5'-CAGGCA **G C A G c T G** TTGG-3'
E-BoxGA	5'-CAGGCA **G C A G a T G** TTGG 3'
E-BoxGT	5'-CAGGCA **G C A G t T G** TTGG-3'
E-Box5'A	5'-CAGGCA **a C A G G T G** TTGG-3'
E-Box5'C	5'-CAGGCA **c C A G G T G** TTGG-3'
E-Box5'T	5'-CAGGCA **t C A G G T G** TTGG-3'
MCK-Mut	5'-CAGGCA **G t g G G c a** TTGG-3'
SP1-WT	5'-GATGCGGTCCCGCCCTCAGC-3'
NOE-Box	5'-CCGAATCTAGGCTATCC-3'

FIGURE 2: Sequences of the oligonucleotides used in the EMSAs. For MCK-S both DNA strands are shown; for all other oligonucleotides only the top strand is shown. The E-box sequences and the base flanking the E-box on the 5′ side of the top strand are shown in boldface; for DNA sequences containing an E-box, changes relative to MCK-S are indicated by lower case lettering.

with the E-box of the MCK enhancer have been well characterized and could therefore serve as a reference point for the comparison of the DNA binding properties of the myogenic and the neurogenic regulators (Lassar et al., 1989; Braun & Arnold, 1991). MASH-1, produced in *E. coli* and purified to apparent homogeneity, was simply mixed with radioactively labeled MCK-S oligonucleotide (Figure 2), and the products of the binding reaction were analyzed by native polyacrylamide gel electrophoresis. Under these conditions only a weak band due to the complex between MASH-1 and DNA was visible. The majority of radioactivity was found in a dark band in the loading well of the gel, indicating that most of the complex between MASH-1 and the oligonucleotide did not enter the gel, probably as a consequence of aggregation and subsequent precipitation. It has been reported previously that MASH-1 binds to an MCK oligonucleotide only very weakly (Johnson et al., 1992). Our results suggested that while MASH-1 could bind to an MCK oligonucleotide, most of the complex formed precipitated and could therefore not be observed as a soluble species giving rise to a band on a polyacrylamide gel. The addition of the nondenaturing zwitterionic detergent CHAPS to the binding reaction kept the protein−DNA complex in solution, allowing the characterization of the DNA binding properties of MASH-1. Binding of MASH-1 to the MCK oligonucleotide was abolished by excess unlabeled MCK-S as a competitor (Figure 3A). Surprisingly, binding was also inhibited in a similar fashion by the addition of an excess of an unlabeled oligonucleotide containing a mutant E-box which has been shown not to be functional (Figure 3A) (Yee & Rigby, 1993). Competition experiments with several unrelated oligonucleotides confirmed these observations.

In titration experiments the apparent dissociation constants (K_D) of the complexes of MASH-1 with a range of oligonucleotides containing E-boxes as well as with oligonucleotides not containing E-box sequences (Table 1) were measured. The concentration of the oligonucleotide was held constant in these measurements, while the protein concentration was successively increased (Figure 3B). The stoichiometry of the MASH-1 complex with the oligonucleotides is 2:1 (see below). The best fit for the isotherms (1) of the binding reactions, describing the dependence of Φ, the fraction of DNA bound, on the concentration of the free protein P (Figure 3C), was found for the assumption of cooperative binding of two monomers to the DNA (Clore et al., 1982):

$$\Phi = K^2[P]^2/(1 + K^2[P]^2) \quad (1)$$

From the isotherms, the protein concentration $[P]_{1/2} = 1/K$, when half of the protein binding sites on the DNA are filled, could be determined and the apparent K_D values calculated from the relation $K_D = ([P]_{1/2})^2$. The values for the complexes of MASH-1 with oligonucleotides containing an E-box sequence ($K_D = 1.4-4.1 \times 10^{-14}$ M^2) were approximately 1 order of magnitude smaller than for the MASH-1 complexes with heterologous oligonucleotides (Table 1).

To prove the stoichiometry of the binding reactions, MASH-1 was mixed with a fragment of MASH-1 comprised of only the BHLH region, and the mixture was heated to 50 °C for 2 min. In an EMSA, one additional complex was observed as well as the DNA complexes of MASH-1 and the BHLH region, indicating that the proteins bind to DNA as dimers (Figure 3D), in accordance with the behavior reported for other BHLH proteins (Sun & Baltimore, 1991).

As all the above experiments were done in the presence of CHAPS in the binding reaction, it was important to determine whether CHAPS was responsible for the low specificity of MASH-1. The addition of bovine serum albumin (BSA) to a final concentration of 0.2% kept the MASH-1−DNA complex in solution up to a protein concentration which corresponded to approximately the square root of the K_D, thereby enabling us to determine approximate dissociation constants for the complexes in the absence of CHAPS by an extrapolation of the data points of the lower half of the sigmoidal curve. The absolute value of the K_D values in these CHAPS-free experiments (Table 1) might represent only an upper limit for the correct value due to some precipitation of the complexes. However, the relative K_D values for the dissociation of the complexes of MASH-1 and different oligonucleotides reflect again the low DNA binding specificity of MASH-1 and indicate that this lack of specificity is an intrinsic property of the protein and not a consequence of the addition of CHAPS.

DNA Binding of the BHLH Region of MASH-1. The high affinity of MASH-1 for DNA suggested that the protein was correctly folded. Nevertheless, the possibility that the low DNA binding specificity displayed by MASH-1 might be due to a partial unfolding of the protein could not be excluded. Given the known modular structure of transcription factors, we determined the DNA binding properties of a peptide comprised of only the BHLH domain of MASH-1 (Figure 1B) as this peptide is small enough to allow us to gain useful information about its structure from CD spectroscopy. In agreement with preliminary data obtained by NMR spectroscopy, the BHLH domain had a much smaller tendency to aggregate than full-length MASH-1, and the

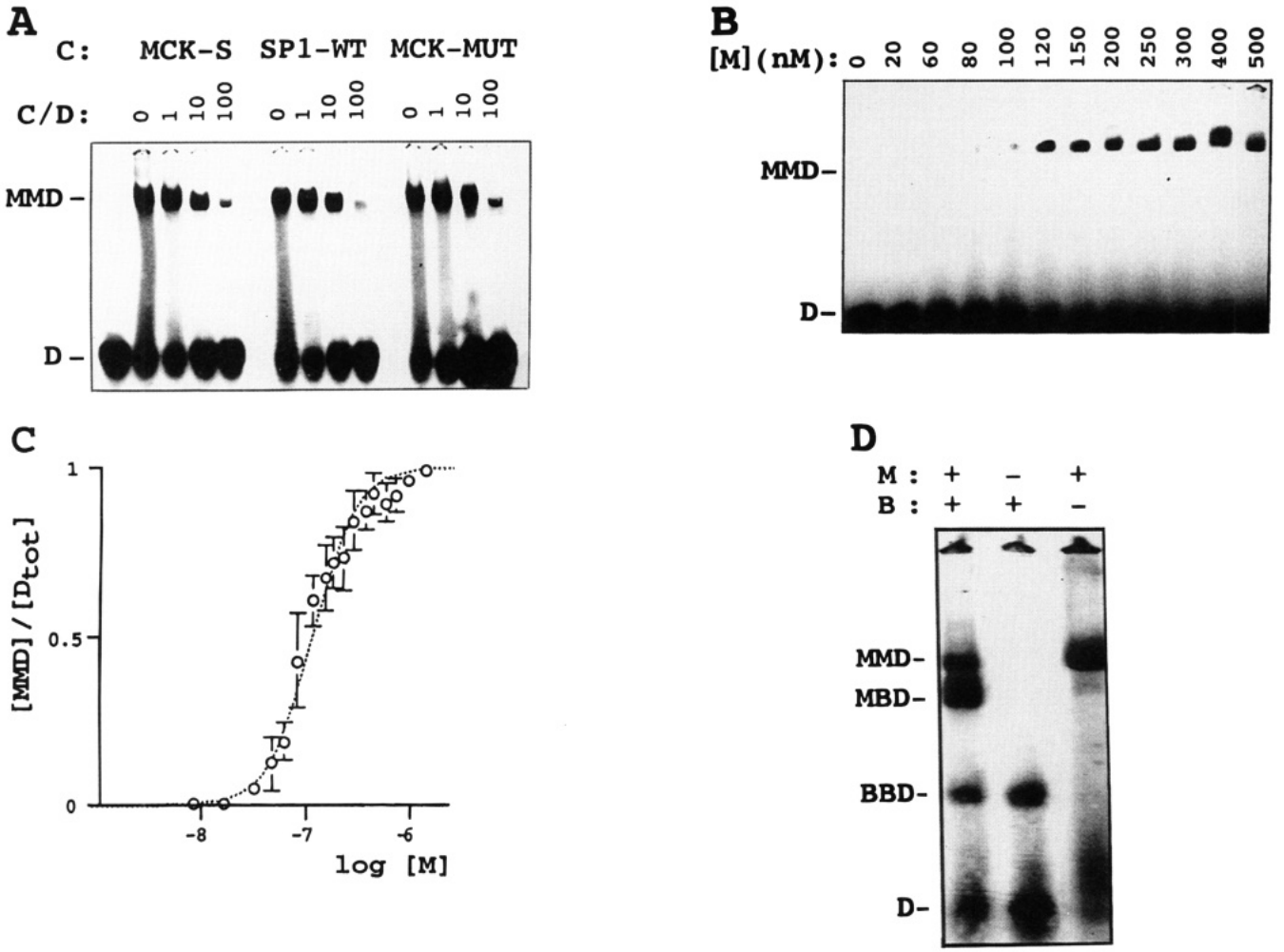

FIGURE 3: Binding of MASH-1 to DNA. (A) Radiolabeled MCK-S oligonucleotide and MASH-1 (except first lane: DNA only) in the presence of increasing concentrations of unlabeled competitor; [MCK-S] = 19 nM; [MASH-1] = 470 nM. (B) Radiolabeled MCK-S in the presence of increasing amounts of MASH-1; [MCK-S] = 19 nM. (C) Fraction of the bound DNA in the MASH-1–MCK-S complex vs the concentration of the free protein for an average of four experiments like the one in (B). The dotted line represents the best fit to eq 1. (D) Mixtures of MASH-1 and BHLH complexed to MCK-S; [MCK-S] = 19 nM; [MASH-1] = 300 nM; [BHLH] = 600 nM. All gels were 4% in acrylamide. All binding reactions contained 4% CHAPS. Abbreviations: D, free DNA; MMD, complex of the MASH-1 homodimer with DNA; C, competitor; C/D, relative amounts of competitor and MCK-S–DNA; M, free MASH-1; B, BHLH domain of MASH-1; BBD, complex of BHLH with DNA; MBD, heterodimer of full-length MASH-1 and BHLH complexed with DNA.

Table 1: Apparent Dissociation Constants for the Protein–DNA Complexes of MASH-1 and E12 and for the Heterodimer of MASH-1 and E12 with and without 4% CHAPS in the Reaction Determined by EMSA[a]

	K_D (M^2)					
	MASH-1		E12		MASH-1/E12	
	+CHAPS	–CHAPS	+CHAPS	–CHAPS	+CHAPS	–CHAPS
MCK-S (5′G)	1.4 (±0.1) E–14	5.5 (±0.6) E–13	2.3 (±0.3) E–15	1.9 (±0.9) E–16	9.9 (±1.5) E–16	7.2 (±1.2) E–17
E-boxCA	1.6 (±0.4) E–14					
E-boxTC	1.7 (±0.3) E–14					
E-boxAC	1.7 (±0.5) E–14					
E-boxGC	1.9 (±0.2) E–14					
E-boxGA	2.1 (±0.1) E–14					
E-boxGT	2.7 (±0.3) E–14					
E-box5′A	1.6 (±0.2) E–14	5.8 (±0.4) E–13	2.8 (±0.3) E–15	2.4 (±0.3) E–16	3.1 (±0.8) E–15	3.8 (±1.9) E–16
E-box5′C	3.9 (±0.4) E–14	9.5 (±0.8) E–13	8.0 (±1.4) E–15	2.0 (±0.4) E–15	2.6 (±0.5) E–15	4.9 (±1.3) E–16
E-box5′T	4.1 (±0.3) E–14	1.4 (±0.2) E–12	1.3 (±0.4) E–14	1.7 (±0.9) E–15	3.3 (±1.2) E–15	6.5 (±1.1) E–16
MCK-Mut	1.4 (±0.4) E–13	1.8 (±0.3) E–12	2.5 (±0.5) E–13	3.6 (±1.3) E–14	7.6 (±2.3) E–15	1.3 (±0.4) E–15
SP1-WT	7.7 (±1.2) E–14	2.0 (±0.2) E–12	1.7 (±0.4) E–13	2.4 (±0.8) E–14	1.4 (±0.4) E–14	1.8 (±0.4) E–15
NOE-box	1.4 (±0.3) E–13	4.2 (±0.5) E–12	3.7 (±0.5) E–13	4.4 (±0.9) E–14	2.6 (±1.4) E–14	1.2 (±0.3) E–15

[a] Standard deviations are given in parentheses.

EMSAs could be performed in the absence of CHAPS or BSA. The apparent K_D for the complex between the BHLH domain of MASH-1 and MCK-S was approximately 1 order of magnitude higher than the K_D for the complex between full-length MASH-1 and MCK-S (Tables 1 and 2), and it was found to be independent of the concentration of CHAPS. However, the BHLH domain displayed virtually no DNA binding specificity under these conditions (Table 2).

Table 2: Effect of the Presence of TFE on the Apparent Dissociation Constant of the MASH-1−BHLH DNA Complex

	K_D (M^2)	
oligonucleotide	−TFE	+20% TFE
MCK-S	1.5 (±0.2) E−13	2.6 (±0.5) E−14
SP1−WT	1.9 (±0.3) E−13	7.2 (±0.7) E−14

The effect of the addition of TFE to the binding reaction was determined because of the known potential of TFE to stabilize the α-helical conformation of peptides (Lehrman et al., 1990; Allemann et al., 1991; Kippen et al., 1994). The affinity of the BHLH domain of MASH-1 for MCK-S was increased by TFE, and this effect reached a maximum when the binding reaction contained approximately 20% TFE (Figure 4A, Table 2). At higher concentrations of TFE the affinity decreased again. TFE also increased the affinity of BHLH for a heterologous oligonucleotide (Table 2). Interestingly, this increase in binding affinity was smaller than with the E-box-containing oligonucleotide. Therefore, the addition of TFE did not lead only to a decrease in the apparent dissociation constants but also to a small, but significant increase in the specificity of the binding reaction, suggesting that TFE stabilizes the native DNA binding conformation of the BHLH domain of MASH-1.

CD Spectroscopy of the BHLH Domain of MASH-1. CD spectroscopy was used to obtain structural information about the BHLH domain of MASH-1. In the absence of DNA, this 69 amino acid polypeptide was largely unfolded at concentrations lower than 5 μM but had considerable α-helical content at concentrations higher than 5 μM. The addition of the MCK oligonucleotide to a 1 μM solution of the BHLH domain induced a change in the CD spectrum (Figure 4B) indicative of the transition in the structure of BHLH from a largely disordered conformation to a mainly α-helical state (for reference, the CD spectrum of the oligonucleotide in the absence of protein is also illustrated in Figure 4B). The binding reaction could be monitored by measuring the intensity of the CD band at 222 nm, which is indicative of the α-helical conformation (Greenfield & Fasman, 1969). In agreement with the results obtained in EMSAs (vide supra), the BHLH domain underwent the same conformational change whether mixed with an oligonucleotide containing or lacking an E-box sequence (Figure 4C). This provided further and independent evidence for the low specificity of DNA binding. The CD spectrum of full-length MASH-1 revealed that its conformation already contained a substantial amount of α-helicity even in the absence of DNA and did not change significantly upon the addition of oligonucleotides, whether or not they contained an E-box

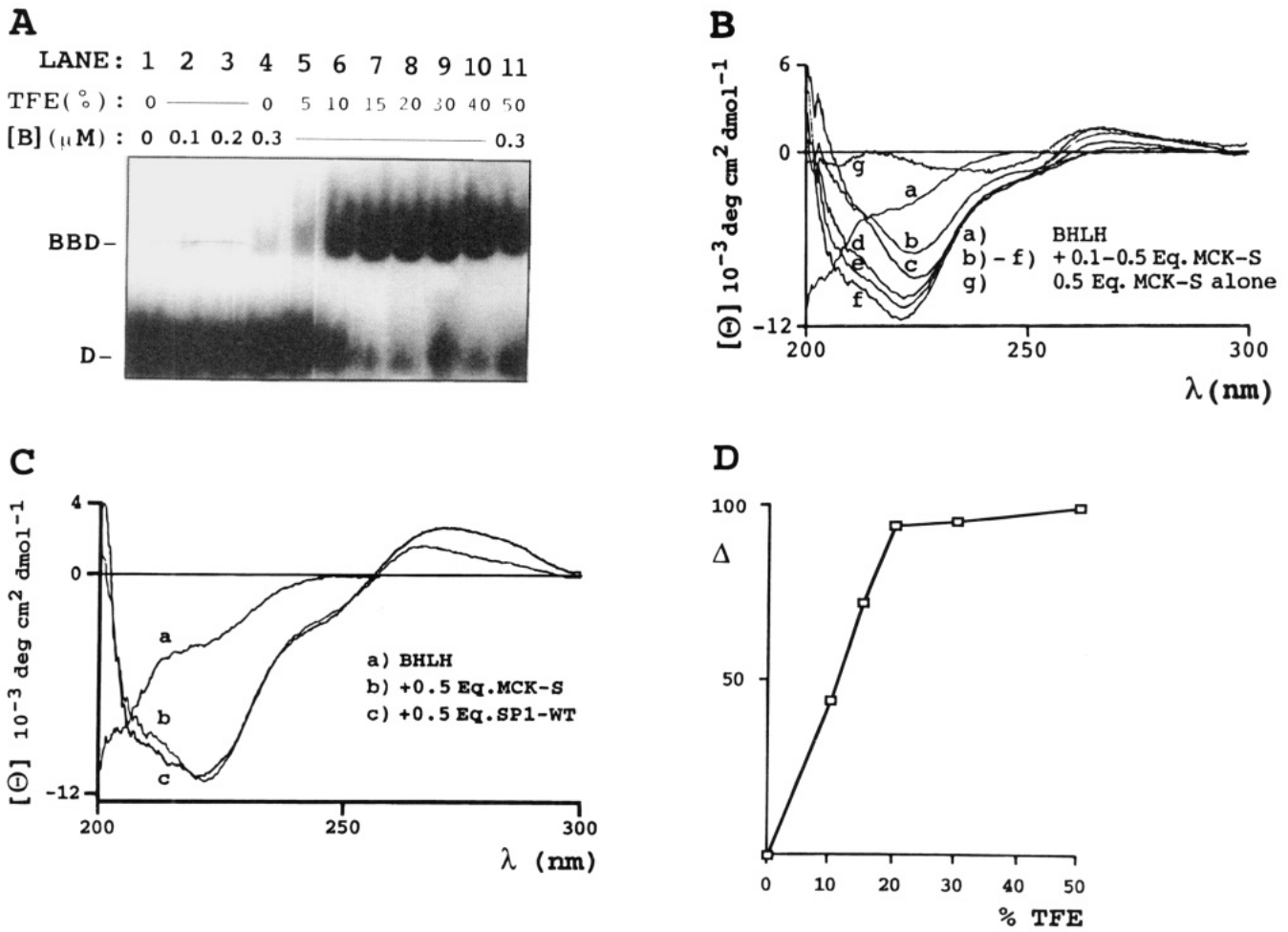

FIGURE 4: Binding of the BHLH domain of MASH-1 to DNA. (A) Radiolabeled MCK-S and BHLH with changing amounts of TFE in the binding reaction; [MCK-S] = 19 nM. (B) CD spectra of BHLH in the presence of increasing amounts of MCK-S oligonucleotide; [BHLH] = 1 μM. Curves: a−f, [MCK-S]/[BHLH] = 0, 0.1, 0.2, 0.3, 0.4, and 0.5; g, [MCK-S] = 0.5 μM, no BHLH present. (C) CD spectra of BHLH in the absence and presence of 0.5 equiv of MCK-S and SP1-WT oligonucleotides; conditions as in (B). (D) Percentage change in $[\Theta]_{222}$, Δ, as a function of the concentration of TFE; [BHLH] = 1 μM. Abbreviations: D, free MCK-S oligonucleotide; B, BHLH domain of MASH-1; BBD, complex of the BHLH domain with MCK-S; [Θ], molar ellipticity; $[\Theta]_{222}$, molar ellipticity at 222 nm.

sequence, suggesting that full-length MASH-1 was already fully folded in the absence of DNA.

The conformational change of the BHLH domain observed upon the addition of DNA could also be induced by the addition of TFE. The content of α-helicity increased already with small concentrations of TFE in the solvent and reached a maximum at approximately 20% TFE (Figure 4D). Θ_{222}, the molar ellipticity at 222 nm of the BHLH domain, was similar whether measured in the presence of 20% TFE or of a stoichiometric amount of an oligonucleotide.

Homodimers of E12 and Heterodimers of MASH-1 and E12. Like MASH-1, E12 showed a tendency to precipitate in aqueous solutions, and 8 M urea and a reducing agent were needed to redissolve the aggregate. However, the formation of the precipitate was very slow, allowing us to determine dissociation constants for the DNA complexes of E12, in both the presence and absence of CHAPS (Table 1). With 4% CHAPS in the binding reaction, E12 homodimers bound to the MCK-S oligonucleotide approximately 10 times better than MASH-1. In detergent-free buffers the binding ($K_D = 1.9 \times 10^{-16}$ M^2) was approximately 3 orders of magnitude stronger for E12 than for MASH-1. This was an upper limit for the increased affinity because of the described inaccuracy of the K_D measurement for MASH-1 under these conditions (vide supra). It should be noted that this high affinity of E12 for E-box sequences is in disagreement with previous reports, describing only very weak binding of E12 to DNA in EMSAs (Sun & Baltimore, 1991). Our result is, however, similar to the K_D of 1.3×10^{-15} M^2 reported for the dissociation of the homodimer of the E47−BHLH domain (Sun & Baltimore, 1991), indicating that the absence of a detectable complex (Sun & Baltimore, 1991) might either be due to the specific fragment used in the EMSAs with E12 or reflect the tendency of E12 to form aggregates. In all the experiments described in this report, E12 was refolded as described in Materials and Methods, and a freshly dialyzed sample was used to avoid problems with aggregation. Table 1 shows that the dissociation constant for the E12 complex with MCK-S in both the presence and absence of CHAPS was approximately 2 orders of magnitude smaller than for the complex with a heterologous oligonucleotide not containing an E-Box sequence. Like MASH-1, E12 showed a preference for a purine flanking C(1) (Figure 2), but the discrimination against a pyrimidine in this position was stronger than in the case of MASH-1.

Simple mixing MASH-1 and E12 in the binding reaction did not allow the heterodimers to form. Therefore, equimolar amounts of the two proteins were mixed and heated to 42 °C for 30 min in the presence of 10 mM DTT followed by slow cooling to room temperature prior to incubation with DNA in the EMSAs. Only one retarded band could be observed (Figure 5) for protein concentrations where only slightly more than 50% binding occurred. When polyclonal antibodies raised against either MASH-1 or E12 were added to the binding reaction prior to electrophoresis, the retarded band was shifted to higher molecular weight by both antibodies, indicating that the complex contained both E12 and MASH-1 (Figure 5). At higher concentrations of protein, a second poorly resolved band of slightly higher molecular weight could be observed due to the formation of the E12 homodimer, indicating that the affinity for DNA of the MASH-1/E12 heterodimer was only slightly higher than that of the E12 homodimer (Table 1). The results of the

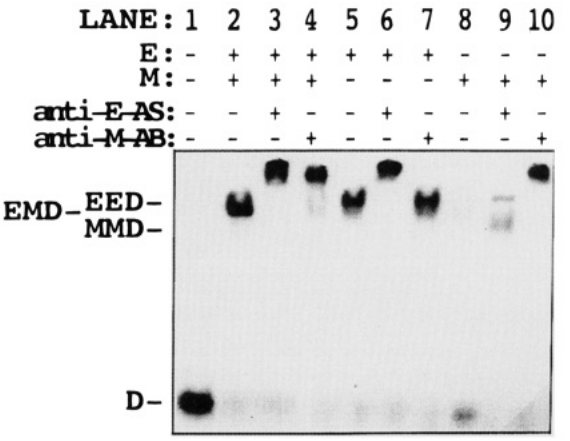

FIGURE 5: Complexes of the homodimers of MASH-1 and E12 and of the heterodimer of MASH-1 and E12 with MCK-S. [MCK-S] = 2.0 nM; [E12] = 0.6 μM (lanes 5−7) and 0.4 μM (lanes 2−4); [MASH-1] = 1 μM (lanes 8−10) and 0.4 μM (lanes 2−4). The weak band of MMD in the absence of antibodies (lane 8) is due to the low solubility of the complex (no CHAPS added to the binding reaction). The intensity of this band was enhanced by the presence of anti-E12 antiserum due to the high protein content of the serum (lane 9). The concentration of E12 antiserum in lane 9 is 5 times higher than in lanes 3 and 6. Abbreviations: D, free MCK-S; E, E12; M, MASH-1; anti-E-AS, antiserum against E12; anti-M-AB, antibody against MASH-1; EED, MMD, and EMD, complexes of MCK-S with the homodimers of E12 and MASH-1 and with the heterodimer of E12 and MASH-1.

evaluation of titration experiments of the heterodimers with various oligonucleotides for protein concentrations, where only complexes between DNA and heterodimer were formed, are shown in Table 1. The dissociation constant for the complex between MCK-S and MASH-1/E12 heterodimers was only slightly smaller than the corresponding value for the E12 homodimer, but binding was 2 orders of magnitude stronger than for the homodimer of MASH-1. The specificity of DNA binding of the heterodimer was reduced by more than 1 order of magnitude, when compared to the specificity of the E12 homodimer. The small amount of specificity of the binding reaction between the MASH1/E12 heterodimer and DNA was comparable to the specificity displayed by the homodimer of MASH-1. It was largely independent of the concentration of CHAPS, analogously to the specificity of E12 and MASH-1 homodimers, thereby providing further evidence that CHAPS was not responsible for the low amount of specificity of these DNA binding proteins. Interestingly, the lower DNA binding specificity of the MASH-1/E12 heterodimers compared to E12 homodimers was due to a substantial stabilization of the complex of the heterodimer with oligonucleotides not containing an E-box and only a 2-fold increase in the binding affinity of MASH-1/E12 heterodimers for E-box sequences (Table 1).

The K_D measurements for the heterodimers showed a preference for a guanine residue flanking the E-box on the 5′ side, analogously to the case of the homodimers of both MASH-1 and E12. Unlike the homodimers, the heterodimers of E12 and MASH-1 showed reduced affinity for oligonucleotides containing an adenine in this position, but they, too, discriminated against pyrimidines. With respect to the preference for the nucleotide 5′ to the E-box, the heterodimers of MASH-1 and E12 and the homodimer of MASH-1 resembled each other, both qualitatively and quantitatively.

DISCUSSION

The isolation and characterization of a large number of sequence-specific DNA binding proteins have shown that many cis-acting DNA sequences that mediate cell-type specific gene expression can interact with not just one transcription factor but with a whole family of structurally related proteins (Lee, 1992; Shirakata et al., 1993). Members of the BHLH family of transcription factors bind in vitro with high affinity to DNA sequences containing an E-box. In vivo, these factors are involved in processes as different as the differentiation of myoblasts or neuroblasts (Weintraub, 1993; Lo et al., 1991). Sequence alignments (Figure 1B) reveal that the BHLH domains of the neurogenic and the myogenic families of proteins show greater sequence similarity of the BHLH domains of the same subfamily than to BHLH sequences that belong to the other subfamily, suggesting that the DNA binding properties of the proteins of the neurogenic and the myogenic subfamilies might be different, providing at least in part an explanation for their different physiological properties. In such a model, the specificity required to control biological processes such as neuronal or muscular differentiation would be achieved by subtle structural changes of the BHLH domain of the individual proteins, thereby inducing specific differences in their DNA binding site preferences (Weintraub et al., 1991; Ma et al., 1994; Fisher & Goding, 1992; Braun & Arnold, 1991).

We have determined the DNA binding properties of MASH-1, a transcription factor involved in neuronal differentiation (Lo et al., 1991), and found that MASH-1 binds to an oligonucleotide containing the E-box sequence of the MCK enhancer as a homodimer with an apparent equilibrium dissociation constant of 1.4×10^{-14} M^2 (Table 1). These measurements were done in the presence of the weak detergent CHAPS because MASH-1 showed a strong tendency to aggregate and precipitate in its absence. The apparent dissociation constant measured without CHAPS (Table 1) represents only an upper limit for the true value and does not reflect an intrinsically low DNA binding affinity of MASH-1 but its tendency to aggregate outside its natural context within the transcription complex, thereby preventing the formation of a soluble DNA−protein complex. The formation of an insoluble precipitate of MASH-1 under detergent-free conditions of the EMSAs might provide an explanation for the apparent discrepancy between our observations and previously published measurements reporting only poor binding of MASH-1 homodimers to the E-box of the MCK enhancer (Johnson et al., 1992). The apparent equilibrium constant measured for the dissociation of the complex of MASH-1 homodimers with MCK-S was comparable to the K_D value reported for the complex of MyoD homodimers with this oligonucleotide (Sun & Baltimore, 1991).

To determine the specificity of DNA binding of MASH-1 and to detect possible differences to the binding specificities of the myogenic regulators, K_D values were measured for the complexes of MASH-1 with a series of oligonucleotides (NCANNTG) with different compositions of the central two base pairs and the base pair 5′ to C(1) (Figure 2 and Table 1). No preference for any nucleotide was found for N(3). For the second internal position, N(4), a slight preference for guanine and cytosine was found. However, the differences in the apparent dissociation constants were small for all substitutions of the internal two bases. The sequence preferences of MASH-1 resembled the preferences reported for the homodimer of MyoD (Sun & Baltimore, 1991; Blackwell & Weintraub, 1990). In an X-ray analysis of cocrystals of the MyoD BHLH domain and an E-box-containing oligonucleotide, no direct contacts between the protein and the two internal bases could be observed (Ma et al., 1994). This lack of direct interactions between the protein and nucleotides N(3) and N(4) might explain why mutations in the central two base pairs left the K_D values largely unchanged. Together with our results, it might also suggest that the clear preference of MyoD homodimers for a GC sequence in the central two base pairs observed in SAAB assays could be the result of an overestimation of the intrinsic DNA binding specificity by that method (Blackwell & Weintraub, 1990). A slightly bigger change in the apparent dissociation constant was found when the base flanking C(1) on the 5′ side was varied. Analogous to the specificity reported for MyoD (Sun & Baltimore, 1991; Blackwell & Weintraub, 1990), MASH-1 discriminated against pyrimidines in this position (Table 1).

According to the crystal structure analysis of MyoD (Ma et al., 1994), most of the amino acids involved in the interaction between helix 1 and helix 2 and in the contacts between the subunits of the dimeric DNA binding complex are either conserved between MASH-1 and MyoD or the change is conservative (Figures 1B and 6). Regions of high sequence variation lie on the outside of the protein in contact with the solvent (Figure 6). There are, however, a few significant differences in the primary sequences of the two proteins. Of the three residues of the BHLH domain of MyoD that make specific contacts with nucleotides in the E-box, only glutamate (118) of MyoD which is hydrogen bonded to C(1) and A(2) (Figure 2) is conserved in MASH-1 [Glu (124)]. Threonine (115) of MyoD, which makes a weak contact to T(5′), is replaced by asparagine (121) of MASH-1. Arginine (120) of MASH-1 can be aligned with arginine (111) of MyoD (Figure 1B), which makes a contact to guanine (6′); it is, however, directly adjacent to Asn (121) and not separated from it by the three residues 112, 113, and 114, like threonine (115) from arginine (111) in MyoD. The neurogenic subfamily of BHLH proteins is the only group with a deletion of three amino acids in this position.

Because of the differences in the primary structure of the BHLH domains of MyoD and MASH-1, the high degree of similarity of the DNA binding preferences displayed by these two proteins was surprising. It could, however, be rationalized post hoc in the following way. Arginine (111) of MyoD makes an hydrogen bond to N7 of G(6′), a water-mediated contact to O(6) of G(6′), and it is hydrogen bonded to the phosphodiester oxygen 5′ of G(6′) (Ma et al., 1994). Arginine (120) of MASH-1 could contact G(6′) in a similar way, even though the contact would be made from the other side of the plane of the guanine base, which lies between these two residues (Figure 7). The water-mediated contact observed in MyoD could be replaced by a direct contact from Arg (120) to O(6) of the guanine base. This argumentation is supported by the observation that, in the X-ray structure of cocrystals of the E12 homodimer and DNA, the asparagine residue, which is in the same position in the α-helix as arginine (120) of MASH-1 (Figure 1B), makes a direct contact to G(6′) (Ellenberger et al., 1994). The same

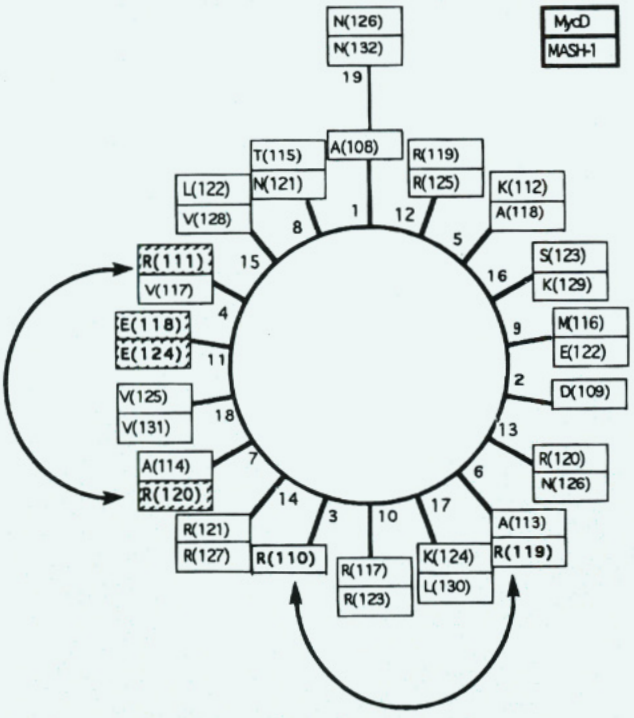

FIGURE 6: Projection of amino acids 108−126 of MyoD (upper box) and 117−132 of MASH-1 (lower box) along the axis of the α-helix formed by the basic domain and helix 1. The two sequences are aligned to give a maximal conservation of the three-dimensional structure of the two proteins. Residues of MyoD that, according to the X-ray structure of the BHLH domain of MyoD and DNA (Ma et al., 1994), make specific contacts with the DNA bases are shown in boldface. Glu (118) of MyoD and Glu (124) of MASH-1 are proposed to make identical contacts to C(1) and A(2) of the DNA. Arg (120) of MASH-1 and Arg (111) of MyoD, which are proposed to make similar contacts to N(7) and O(6) and G(6′), and Arg (119) and Arg (110), which might make a similar phosphate contact, are connected by arrows.

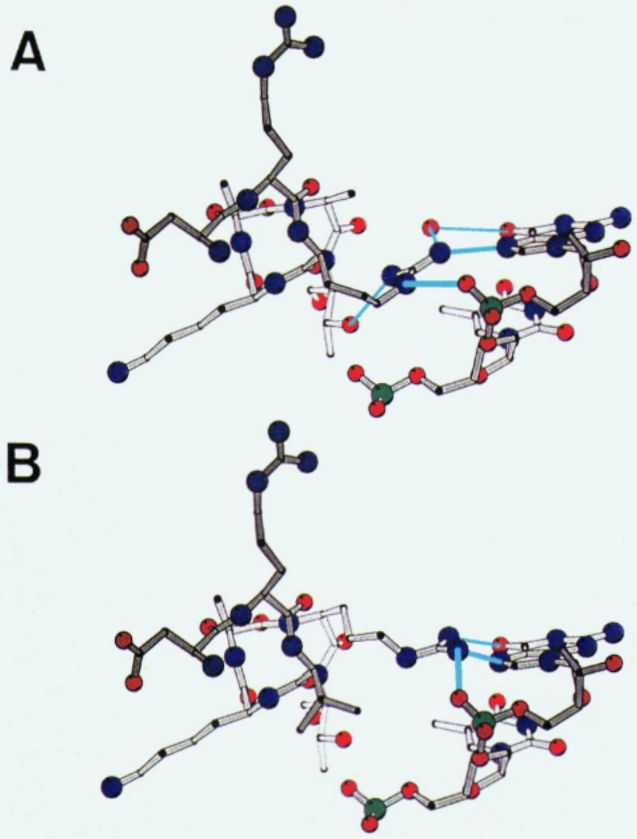

FIGURE 7: Contacts of MyoD and MASH-1 with guanine (6′) of an E-box sequence. (A) Diagram showing the structure of the BHLH domain of MyoD from amino acid Asp (109) to Thr (115) and the key contacts of Arg (111) with guanine (6′), according to the published crystal structure (Ma et al., 1994). Note that the contact of Arg (111) to O(6) of guanine is mediated by a water molecule. (B) Proposed structure for the α-helical segment in (A), in which Arg (111) and Ala (114) were replaced by Val and Arg, respectively, and the proposed contacts to guanine (6′). For the two amino acid residues, which were changed in the peptide of (B), the orientation of all C(α)−C(β) bonds was left unchanged, all torsion angles are within the favorable range, and no steric hindrance was introduced. No changes of the coordinates of atoms not belonging to the side chains of these two residues were allowed. The pictures were generated on a Power Macintosh 7100 computer, using the program MoMo (Dobler, 1992). For the final display the program Molscript (Kraulis, 1991) was used. Phosphorus atoms are green, nitrogen atoms blue, and oxygen atoms red. Carbon atoms are in gray, and the shade of gray is varied to give depth to the picture. Relevant hydrogen bonds are indicated by a light blue line.

mechanism could also allow arginine (119) of MASH-1 to interact with the phosphate backbone in a way similar to that of arginine (110) of MyoD, despite the separation of these two residues by almost one full turn in the putative α-helix of the superimposed basic domains (Figure 6). Only for the weak hydrophobic contact of the γ-carbon of threonine (115) of MyoD with the methyl group of T(5′) (Ma et al., 1994), no possible replacement could be found. However, the main function of this threonine in MyoD is to keep Arg (111) in contact with G(6′) through a hydrogen bond between the hydroxyl group of Thr and the side chain of arginine (111). Without this hydrogen bond, Arg (111) might swing away from the guanine base and solely make a contact with the backbone phosphate (Figure 7). Such a function of threonine (115) is supported by the observation that, in the cocrystals of DNA and E47, in which threonine is replaced by asparagine, the E47 arginine, which corresponds to Arg (111) of MyoD, makes only a contact to the DNA backbone and no specific contacts to the bases (Ellenberger et al., 1994). Therefore, it is plausible to propose that no such threonine residue is required in MASH-1, as arginine (120), which we propose to make specific contacts to G(6′), is in a different helix position than Arg (111) of MyoD.

It should be kept in mind that the explanation provided above for the similar sequence specificities of MASH-1 and E12 can only be adequate qualitatively. The altered geometry of the interactions between DNA and the BHLH domains will affect the strength of the observed sequence preferences, and therefore, slightly different contacts will be optimal in the two geometries. Consequently, the extent of the sequence preference will be different for different BHLH proteins (Table 1).

The observation that the preference of both MASH-1 and E12 homodimers for a certain E-box results only in a change of the apparent dissociation constant of less than a factor of 10 (Table 1) suggests that the specificity of transcription of the target genes of neurogenic regulators like MASH-1 cannot be achieved by a mechanism which is solely governed by differential binding to different E-box sequences. This conclusion is supported by recently published results from experiments with transgenic mice (Yee & Rigby, 1993). The proximal E-box sequence (GCAGTTG) of the myogenin promoter (most likely a target for the myogenic BHLH protein Myf-5) and a binding site for RSRF (*r*elated to *s*erum

response factor) were described to be sufficient to recapitulate the proper expression pattern of the myogenin gene. A mutation of the proximal E-box sequence, which abolished DNA binding, also prevented expression of the transgene. However, when an E-box with the sequence TCACATG was added to the construct containing the mutated proximal E-box and an RSRF site, the expression pattern was indistinguishable from the wild-type pattern.

The small differences for the dissociation constants for different E-box sequences notwithstanding, it was nevertheless surprising to find that homodimers of MASH-1 bound with only 10 times reduced affinity to sequences not containing an E-box at all (Table 1). This very small specificity was not a consequence of the addition of the detergent CHAPS to the binding reaction, since it was also observed in its absence (Table 1). In the case of E12, the specificity of DNA binding was more than 1 order of magnitude higher (Table 1).

For the BHLH domain of MASH-1 alone, the specificity of DNA binding was even smaller (Table 2). The dissociation constants measured for the BHLH domain were approximately 1 order of magnitude smaller than for full-length MASH-1. CD spectroscopy showed that in the absence of DNA the BHLH domain was largely unfolded at the concentrations of the EMSAs but underwent at transition to a mainly α-helical conformation upon binding to DNA, independently of the nature of the oligonucleotide (Figure 4B,C). Interestingly, the same transition to an α-helical conformation could also be induced by the addition of 20% TFE to the binding reaction (Figure 4D). In addition, 20% TFE also increased the affinity of BHLH for DNA, which in the presence of the cosolvent was similar to that measured for full-length MASH-1, suggesting that TFE stabilized the native DNA binding conformation of MASH-1. The observation that in the presence of 20% TFE the specificity of DNA binding of the BHLH domain was similar to the specificity of full-length MASH-1 (Tables 1 and 2) implied that the low DNA binding specificity of MASH-1 was not a consequence of a partially unfolded conformation of the BHLH domain but an intrinsic property of this protein.

The myogenic proteins can form heterodimers with the ubiquitous proteins, E12 and E47, and the heterodimers bind to DNA with at least 10 times higher affinity than either of the homodimers (Sun & Baltimore, 1991). SAAB has shown that the hetero- and homodimers bind to DNA with slightly different sequence preferences (Blackwell & Weintraub, 1990). Strong evidence has been provided that in vivo the heterodimers of the myogenic regulators and E12/47 are the functionally active species (Lassar et al., 1991), and it has been suggested that the biologically active species of MASH-1 could also be a heterodimer with E12 or E47 (Lo et al., 1991; Johnson et al., 1992). Our results show that, despite the higher affinity for DNA as compared with the MASH-1 and E12 homodimer, the heterodimer between MASH-1 and E12 displayed significantly less DNA binding specificity than the E12 homodimer. Interestingly, this effect was mainly a consequence of improved binding of DNA sequences not containing an E-Box and only to a lesser extent because of improved binding of E-box sequences (Table 1). The DNA binding preferences of the heterodimer paralleled the preferences of the MASH-1 homodimer even though they were more pronounced by a factor of 2 in the presence of CHAPS and by a factor of 4 in its absence. Therefore, heterodimerization with E12 did not significantly enhance the specificity of DNA binding of MASH-1, and except for slightly tighter binding to DNA, no advantage of the heterodimer over the homodimer could be detected in EMSAs.

The neurogenic subfamily of transcription factors containing BHLH motifs might represent the extreme case of low specificity of DNA binding in a family of transcription factors with generally only moderate sequence preferences. The conclusion reached for MASH-1 from the above results, that the specificity necessary to ensure cell-type specific expression must to a large extent come from posttranslational modifications of MASH-1 or from interactions with other components of the transcriptional machinery, might only be true partly for other BHLH proteins, such as the myogenic regulators. However, it suggests that combinatorial interactions with other transcription factors are of greater importance for the specificity of gene expression mediated by BHLH proteins that previously suspected.

ACKNOWLEDGMENT

We thank Steven A. Benner for his support and for the use of his laboratory; David J. Anderson for providing the cDNAs of MASH-1 and of E12; Carl O. Pabo for supplying the coordinates of the crystal structure of the DNA complex of the MyoD BHLH domain; Martin Egli and Stefan Portmann for making Figure 7; and Lesley A. Tannahill, Martin Egli, and Andreas Künne for critical reading of the manuscript and for helpful discussions.

REFERENCES

Allemann, R. K., Presnell, S. R., & Benner, S. A. (1991) *Protein Eng. 4*, 831−835.
Blackwell, T. K., & Weintraub, H. (1990) *Science 250*, 1104−1110.
Braun, T., & Arnold, H. H. (1991) *Nucleic Acids Res. 19*, 5645−5651.
Buonanno, A., Apone, L., Morasso, M. I., Beers, R., Brenner, H. R., & Eftimie, R. (1992) *Nucleic Acids Res. 20*, 539−544.
Buskin, J. N., & Hauschka, S. D. (1989) *Mol. Cell. Biol. 9*, 2627−2640.
Campuzano, S., & Modolell, J. (1992) *Trends Genet. 8*, 202−208.
Clore, G. M., Gronenborn, A. M., & Davies, R. W. (1982) *J. Mol. Biol. 155*, 447−466.
Davis, R. L., & Weintraub, H. (1992) *Science 256*, 1027−1030.
Davis, R. L., Weintraub, H., & Lassar, A. B. (1987) *Cell 51*, 987−1000.
Davis, R. L., Cheng, P.-F., Lassar, A. B., & Weintraub, H. (1990) *Cell 60*, 733−746.
Dobler, M. (1992) *MoMo (color version 1.4)* ETH-Zurich.
Edmondson, D. G., & Olson, E. N. (1989) *Genes Dev. 3*, 628−640.
Ellenberger, T., Fass, D., Arnaud, M., & Harrison, S. C. (1994) *Genes Dev. 8*, 970−980.
Fisher, F., & Goding, C. R. (1992) *EMBO J. 11*, 4103−4109.
German, M. S., Blanar, M. A., Nelson, C., Moss, L. G., & Rutter, W. J. (1991) *Mol. Endocrinol. 5*, 292−299.
Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108−4115.
Hasty, P., Bradley, A., Morris, J. H., Edmondson, D. G., Venuti, J. M., Olson, E. N., & Klein, W. H. (1993) *Nature (London) 364*, 501−506.
Johnson, J. E., Birren, S. J., & Anderson, D. J. (1990) *Nature (London) 346*, 858−861.
Johnson, J. E., Birren, S. J., Saito, T., & Anderson, D. J. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3596−3600.
Kippen, A. D., Sancho, J., & Fersht, A. R. (1994) *Biochemistry 33*, 3778−3786.

Kraulis, R. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Lassar, A. D., Buskin, J. N., Lockshorn, D., Davis, R. L., Apone, S., Hauschka, S. D., & Weintraub, H. (1989) *Cell 58*, 823−831.
Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronona, A., Baltimore, D., & Weintraub, H. (1991) *Cell 66*, 305−315.
Lee, K. A. W. (1992) *J. Cell. Biol. 103*, 9−14.
Lehrman, S. R., Tuls, J. L., & Lund, M. (1990) *Biochemistry 29*, 5590−5596.
Lo, L. C., Johnson, J. E., Wuenschell, C. W., Saito, T., & Anderson, D. J. (1991) *Genes Dev. 5*, 1524−1537.
Ma, P. C. M., Rould, M. A., Weintraub, H., & Pabo, C. O. (1994) *Cell 77*, 451−459.
Miner, J. H., & Wold, B. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1089−1093.
Murre, C., McCaw, P. S., & Baltimore, D. (1989) *Cell 56*, 777−783.
Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S., Nanoka, I., & Nabeshima, Y. (1993) *Nature (London) 364*, 532−535.
Olson, E. N. (1990) *Genes Dev. 4*, 1454−1461.
Paterson, B. M., Walldorf, U., Eldridge, J., Dübendorfer, A., Fratsch, M., & Gehring, W. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3782−3786.
Rudnicki, M. A., Schnegelsberg, P. N. J., Stead, R. H., Braun, T., Arnols, H. H., & Jaenisch, R. (1993) *Cell 75*, 1351−1359.
Shirakata, M., Friedman, F. K., Wei, Q., & Paterson, B. M. (1993) *Genes Dev. 7*, 2456−2470.
Studier, F. W. (1991) *J. Mol. Biol. 219*, 37−44.
Studier, F. W., & Moffat, B. A. (1986) *J. Mol. Biol. 189*, 113−130.
Sun, X.-H., & Baltimore, D. (1991) *Cell 64*, 459−470.
Weintraub, H. (1993) *Cell 75*, 1241−1244.
Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B., & Miller, A. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5434−5438.
Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y., & Lassar, A. (1991) *Science 251*, 761−766.
Wetlaufer, D. R. (1962) *Adv. Protein Chem. 17*, 303−390.
Wright, W. E., Binder, M., & Funk, W. (1991) *Mol. Cell. Biol. 11*, 4104−4110.
Yee, S.-P., & Rigby, P. W. J. (1993) *Genes Dev. 7*, 1277−1289.

BI950913J