1408–1413 *Nucleic Acids Research, 1998, Vol. 26, No. 6*

Single chain dimers of MASH-1 bind DNA with enhanced affinity

Martin Sieber and Rudolf K. Allemann*

Laboratory for Organic Chemistry, Department of Chemistry, ETH-Zürich, Universitätstrasse 16, CH-8092 Zurich, Switzerland

Received December 5, 1997; Revised and Accepted January 27, 1998

ABSTRACT

By designing recombinant genes containing tandem copies of the coding region of the BHLH domain of MASH-1 (MASH-BHLH) with intervening DNA sequences encoding linker sequences of 8 or 17 amino acids, the two subunits of the MASH dimer have been connected to form the single chain dimers MM8 and MM17. Despite the long and flexible linkers which connect the C-terminus of the first BHLH subunit to the N-terminus of the second, a distance of ~55 Å, the single chain dimers could be produced in *Escherichia coli* at high levels. MM8 and MM17 were monomeric and no 'cross-folding' of the subunits was observed. CD spectroscopy revealed that, like wild-type MASH-BHLH, MM8 and MM17 adopt only partly folded structures in the absence of DNA, but undergo a folding transition to a mainly α-helical conformation on DNA binding. Titrations by electrophoretic mobility shift assays revealed that the affinity of the single chain dimers for E box-containing DNA sequences was increased ~10-fold when compared with wild-type MASH-BHLH. On the other hand, the affinity for heterologous DNA sequences was increased only 5-fold. Therefore, the introduction of the peptide linker led to a 4-fold increase in DNA binding specificity from –0.14 to –0.57 kcal/mol.

INTRODUCTION

The basic helix–loop–helix family of eukaryotic transcription factors relies on a simple structural motif for sequence-specific DNA recognition. The DNA binding activity of these proteins is confined to ~60 amino acids, named the basic helix–loop–helix (BHLH) domain (1–3; Fig. 1A). The BHLH domain comprises two regions of distinct function in DNA recognition, the helix–loop–helix domain, which mediates dimerization, and the basic region, which contacts the DNA through direct interactions with the phosphate backbone and the nucleobases (Fig. 1B; 4,5). Results from circular dichroism (CD) spectroscopy showed that in the absence of DNA BHLH proteins can form stable dimers, which are found in a concentration-dependent equilibrium with the monomer (6,7). Dimerization is accompanied by a folding transition from the largely unfolded monomer to a mainly α-helical dimer, in which helices 1 and 2 are separated through a loop of ~8 amino acids. The same transition can be induced by addition of DNA, even at concentrations where the BHLH domain alone is mainly unfolded (7–9). NMR spectroscopy and ITC experiments have shown that in the absence of DNA the basic region remains unfolded, even at concentrations where the dimer is the predominant species (10,11). However, upon DNA binding the basic region also adopts an α-helical conformation. The crystal structure analyses of the DNA complexes of the BHLH proteins E47 and MyoD revealed that the basic region is simply the N-terminal end of helix 1 and that helices 1 and 2 form the tightly packed core of the dimers (Fig. 1B; 4,5).

Surprisingly, the DNA binding specificity displayed by BHLH proteins was found to be small. The BHLH domain of MASH-1 (MASH-BHLH) binds to E box-containing DNA with only marginally higher affinity than to heterologous sequences (8,9). However, covalently linking the subunits of MASH-BHLH through the introduction of a disulfide bond at the C-terminal end of helix 2 increased the DNA binding specificity ~5-fold (7). The linkage enforced the close proximity of the two helix 2 regions of the individual subunits. In sharp contrast to wild-type MASH-BHLH, the crosslinked 'dimer' was found to be stably folded, even in the absence of DNA.

The subunits of many multimeric protein complexes can be connected through the introduction of covalent linkers. For example, the α- and β-subunits of glycyl-tRNA synthetase could be fused via a short peptide linker, creating a fully active single chain protein (12). Other examples include CuZn superoxide dismutase (13), avian retroviral proteases (14), the RNA binding protein ROP (15), the sweet tasting peptide monellin (16), single chain antibodies (17–19) and both the 434 and arc repressors (20,21). 'Single chain multimers' provide an approach to the creation of hybrid proteins with novel properties, such as specificity or activity. Single chain fusions can be displayed on filamentous phages and novel specificities and affinities can be selected for from large repertoires of mutant proteins (17,22–24). Random mutagenesis and *in vitro* selection by phage display has been used to create variants of Zn finger proteins with altered DNA binding properties (25,26). These experiments were greatly facilitated by the monomeric nature of these transcription factors.

Here we describe the construction and investigate the conformational and DNA recognition properties of 'single chain dimers' of MASH-BHLH in which the C-terminus of one BHLH subunit is attached to the N-terminus of the second through peptide linkers of varying length (Fig. 1B). Independent of the linker used, the 'single chain dimers' bound to DNA with significantly enhanced affinity and specificity. Unlike disulfide-linked MASH-BHLH, the single chain dimers did not adopt a fully folded structure in the absence of DNA, but underwent a folding transition on DNA binding. These results show that linking

*To whom correspondence should be addressed. Tel: +41 1 632 2895; Fax: +41 1 632 1170; Email: allemann@chem.ethz.ch

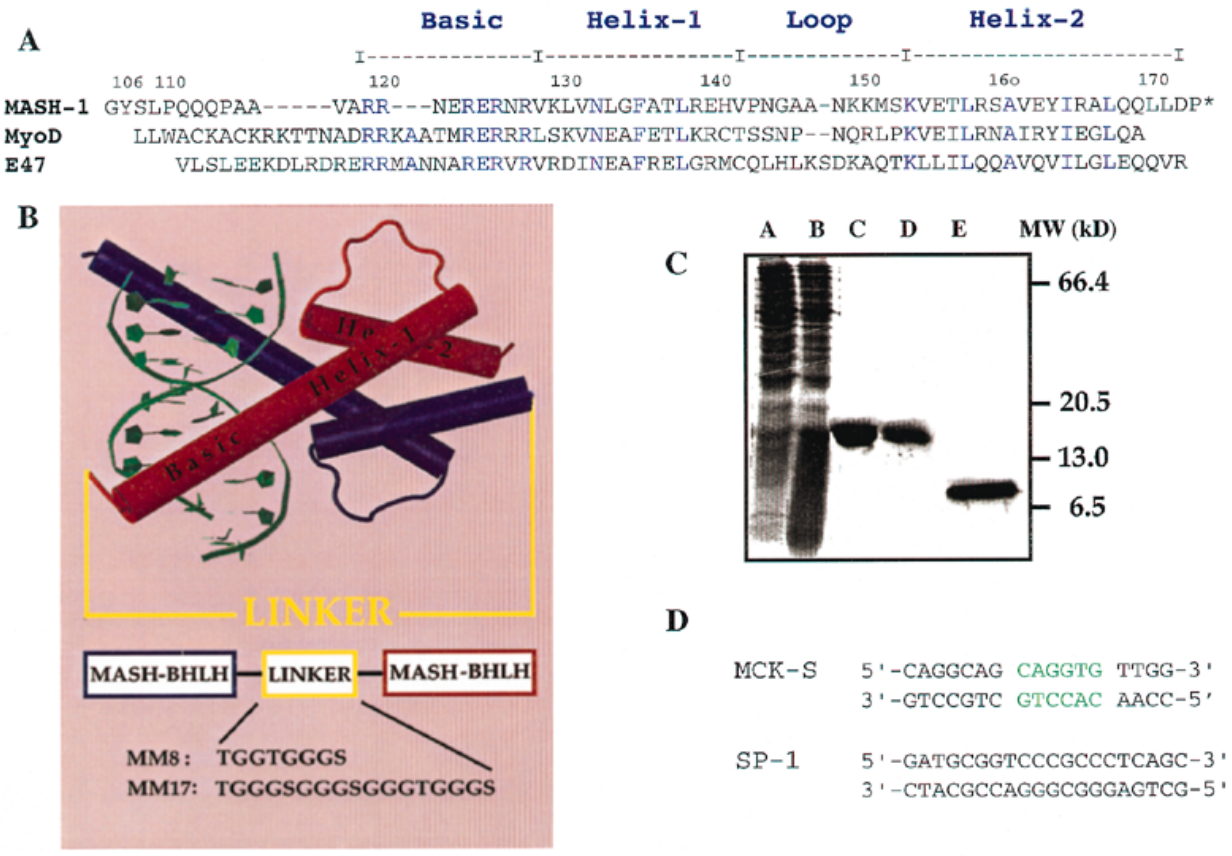

Figure 1. (**A**) Alignment of the BHLH domains of MASH-1 (MASH-BHLH) (8), MyoD (35) and E47 (36). The numbering system corresponds to full-length MASH-1. The proline marked * is a cloning artefact and is not part of the MASH-1 cDNA. MM8 and MM17 contain this proline at their C-termini. The positions of the basic region, helices 1 and 2 and the loop are based on the co-crystal structures of MyoD (4) and E47 (5) with DNA. (**B**) Sketch of the DNA complexes of MM8 and MM17. The first BHLH subunit is coloured blue and the second subunit red. The DNA is coloured green. The linker region connecting the C-terminus of the first subunit to the N-terminus of the second is indicated in yellow. The respective linker sequences are given in the one letter code. The programs VMD and Raster 3D (37,38) were used to create this display from the coordinates of the DNA complex of MyoD (4). (**C**) SDS–PAGE of crude extracts of *E.coli* cells harbouring an expression plasmid for MM17 just before (lane A) and 2 h after induction of expression (lane B). Lanes C–E, purified proteins MM8, MM17 and MASH-BHLH (8). Mobilities of molecular weight marks (MW) are given in kDa. (**D**) Sequences of the oligonucleotides used in this study. Both strands are shown and the E box of MCK-S is highlighted in green.

the subunits of MASH-BHLH through a peptide linker does not significantly alter the folding and DNA recognition properties of MASH-BHLH. The increased affinity and specificity are most likely due to a linker-induced reduction in the conformational freedom of the basic region in the disordered state.

MATERIALS AND METHODS

Construction of expression plasmids for MM8 and MM17

The gene encoding MM8 was constructed in three steps. Plasmid pJGetMASH-BHLH, which contains a fragment of the MASH-1 cDNA coding for the BHLH domain from G(106) to D(172) (8), was digested with restriction enzymes *Pst*I and *Bam*HI. The resulting vector fragment was ligated with a cassette with sequence

```
  5′-g cag ctg ctg ACC GGT GGT ACC GGg
ac gtc gtc gac gac TGG CCA CCA TGG CCc cta g-5′
```

resulting in plasmid pJgetMABlink1 (lower case letters indicate bases from the coding region of the *MASH* gene). In a second step a cassette (coding for the second half of the linker) with sequence

```
5′-T ATG GGT ACC GGG GGT GGA AGT AT
     AC CCA TGG CCC CCA CCT TCA TAA t-5′
```

was inserted into the *Nde*I site of pJGetMASH-BHLH to give plasmid pJGetMABlink2. In the final step the *Kpn*I–*Bam*HI fragment of the insert in pJGetMABlink2 was inserted between the *Kpn*I and *Bam*H sites of pJGetMABlink1 to yield pJGetMM8.

To construct the expression plasmid for MM17, pJGetMM8 digested with *Age*I and *Kpn*I and the DNA sequence coding for the additional amino acids of the linker were inserted through ligation with the following double-stranded oligonucleotide

```
5′-CC GGT GGA GGT AGT GGT GGC GGG TCA GGT GGA GGT AC
     A CCT CCA TCA CCA CCG CCC AGT CCA CCT C-5′
```

The DNA sequence of all constructs was verified using the dideoxy sequencing method (27).

BL21(DE3)pLysS cells containing the MM8 or MM17 expression plasmids were grown at 37°C on LB medium with 100 mg/l ampicillin and 50 mg/l chloramphenicol until the OD_{600} reached 0.4. Then IPTG was added to a final concentration of 1 mM. Cells were harvested 3 h after induction by centrifugation and pellets were frozen at –20°C.

Purification of MM8 and MM17

MM8 and MM17 were purified essentially as described for the BHLH region of MASH-1 and for the MASH mutant MASH-GGC

(7,8). The purified proteins were homogeneous as judged by SDS–PAGE and cation exchange chromatography on a Resource-S (Pharmacia) HPLC column. MALDI-TOF mass spectrometry showed molecular masses of 16 001 and 16 581 for MM8 and MM17 respectively, which corresponded well with the calculated masses of 15 977 and 16 580 for the single chain dimers without their N-terminal methionines. Sequencing by Edman degradation gave the correct N-terminal sequences and confirmed that the N-terminal methionine had been removed proteolytically. Protein concentrations were determined by measuring the UV absorption at 215 and 220 nm (28). The yields for the preparations were ~4 mg purified protein/l culture.

CD spectroscopy

CD spectra were measured using a Jasco J600 spectropolarimeter. The buffer was 1 mM Tris–HCl, pH 7. Spectra were measured for a concentration range of 100 nM–5 μM. For DNA binding experiments the protein concentrations were 0.5 μM for MM8 and MM17; for MASH-BHLH a concentration of 1 μM was used.

Oligonucleotides

Oligonucleotides were purchased from Microsynth, desalted on Sephadex and precipitated with ethanol. Double-stranded MCK-S oligonucleotide, containing a central E box sequence and SP-1 oligonucleotide, were used as specific and heterologous DNA probes respectively (Fig. 1D). Single-stranded oligonucleotides were labelled with [γ-^{32}P]ATP (Amersham) in the presence of T4 polynucleotide kinase (NEB) and complementary strands (10% excess) were annealed by heat denaturation followed by slow cooling to room temperature.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSA) were performed as previously described (7,8). Bacterially expressed proteins were serially diluted into EMSA buffer (50 mM Tris–HCl, pH 7.9, 6 mM $MgCl_2$, 40 mM ammonium sulphate, 0.2 mM EDTA, 1 mM DTT, 5% glycerol). This solution was incubated in the presence of 10 nM labelled oligonucleotide for 10 min at room temperature. Samples were applied to 4% polyacrylamide gels in 0.9× TAE, pH 7.9. After electrophoresis the gels were dried and exposed to Kodak X-OMAT-S film at –70°C. Quantitative data were obtained with a Packard Instantimager using system software. The fraction Φ of DNA bound was determined as the activity of the retarded band (corresponding to the protein–DNA complex) divided by the sum of the activities of the retarded and unretarded (corresponding to the free DNA) bands. Plotting Φ against the concentration of unbound protein allowed determination of the concentration $[P]_{1/2}$ at which half of the protein binding sites were filled (8). The best fit for DNA binding of 'single chain dimers' to the binding isotherm (1)

$$\Phi = 1/(1 + [P]_{1/2}/[P]^n) \quad (1)$$

was obtained for $n = 1$.

RESULTS AND DISCUSSION

Design, expression and purification of 'single chain dimers' of MASH-BHLH

The association reaction between BHLH proteins and DNA is characterized through the energetic coupling of protein folding, dimerization and DNA binding (7,8,29,30). Data from CD and nuclear magnetic resonance spectroscopy revealed that in the absence of DNA the helix–loop–helix domain can form a stably folded dimer which is found in a concentration-dependent equilibrium with the unstructured monomer with dimerization constants between 1 and 50 μM (6–8,10). However, at the concentrations where half maximal DNA binding occurs (10–500 nM) BHLH proteins are largely unfolded monomers in solution. Folding and dimerization are induced upon DNA binding. Therefore, the favourable free energy of the association reaction is reduced, because some energy must be spent on dimerization and folding at concentrations where dimerization is unfavourable. We have shown that linking the subunits of MASH-BHLH through a disulfide bond not only obviated the requirement for dimerization, but also induced the protein to adopt the folded conformation even in the absence of DNA (7). Here we tested the hypothesis that linking the C-terminus of the first BHLH subunit to the N-terminus of the second through a peptide linker should result in increased DNA binding activity without significantly altering the conformational properties of the protein.

According to crystal structure analyses of the DNA complexes of MyoD and E47 the shortest path between the C- and N-termini of the two protein subunits is ~55 Å (Fig. 1B; 4,5). Therefore, a linker of 17 amino acid residues seemed sufficient to connect the two monomers, resulting in the 'single chain dimer' MM17 (Fig. 1B). Since the primary sequence of MASH-BHLH suggested that the nine N-terminal amino acids might not adopt an α-helical conformation, we also constructed MM8, in which two MASH-BHLH domains are connected through an eight residue linker. Successful construction of an active 'single chain dimer' depends on a linker that neither interferes with folding and association of the two BHLH domains nor reduces stability and recognition properties of MASH-BHLH. Many surface loops in natural proteins consist of glycine, threonine and serine residues and we chose these residues for our linkers in order to maximize both flexibility and solubility (Fig. 1B).

The single chain dimers MM8 and MM17 were produced in *E.coli* and purified to apparent homogeneity (Fig. 1C). They could be expressed at levels similar to wild-type MASH-BHLH. The yields of purified proteins were similar to wild-type levels, indicating that proteolytic degradation of the flexible linkers was not a problem.

CD spectroscopy of the single chain dimers MM8 and MM17

CD spectroscopy was used to obtain structural information about MM8 and MM17. The CD spectrum of a 1 μM solution of wild-type MASH-BHLH revealed that ~25% of the amino acids were in an α-helical conformation (Fig. 2A; 31,32). Even though the amount of α-helical structure was higher in the single chain dimers (~38%), a significant portion of the peptides remained unstructured (Fig. 2A). This is in sharp contrast to the behaviour of the MASH mutant MASH-GGC, in which under oxidizing conditions the BHLH subunits are held together through a disulfide bond at the C-terminal end of helix 2 (7). Oxidized

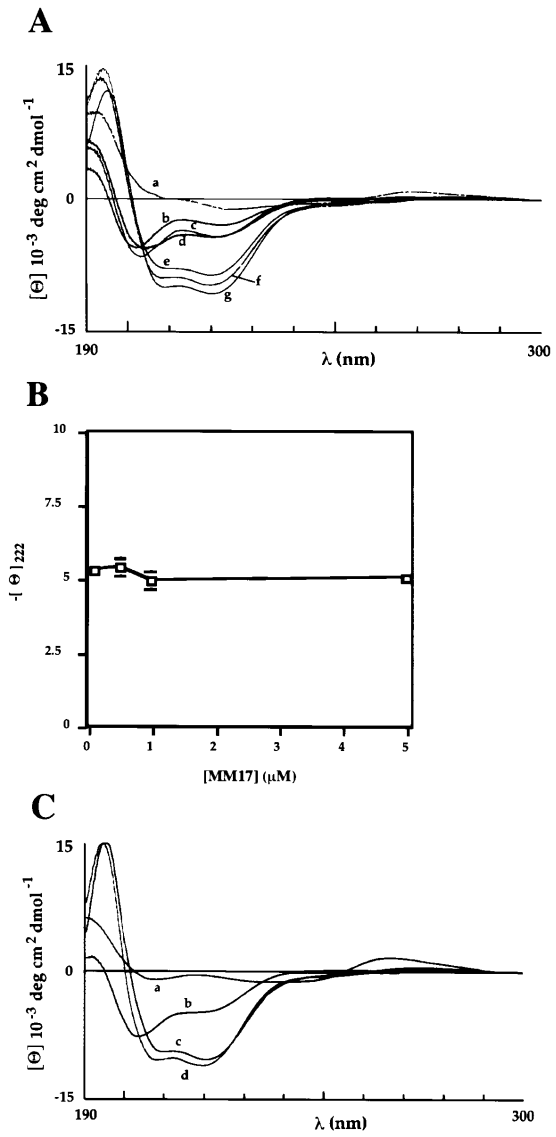

Figure 2. (**A**) Circular dichroism spectra of MASH-BHLH (b), MM8 (d) and MM17 (c) and of the MCK-S complexes of MASH-BHLH (f), MM8 (e) and MM17 (g); (a) CD spectrum of 0.5 μM MCK-S (no protein present). [MASH-BHLH] 1 μM; [MM8] and [MM17] 0.5 μM (corresponding to 1 μM BHLH equivalents); [MCK-S] 0.5 μM. (**B**) Molar ellipticity at 222 nm, $[\theta]_{222}$, from CD spectra of MM17 as a function of the concentration of MM17. Concentrations of MM17 were 0.1, 0.5, 1 and 5 μM. The errors obtained from multiple measurements under identical conditions are indicated. (**C**) CD spectra of MM8 in the presence and absence of oligonucleotides MCK-S and SP-1. (a) MCK-S alone; (b) MM8 alone; (c) and (d) MM8 plus 1 equiv. SP-1 (c) or MCK-S (d). [MM8] 0.5 μM; [MCK-S] 0.5 μM; [SP-1] 0.5 μM.

MASH-GGC was stably folded and mainly α-helical. The disulfide linkage keeps two segments of the BHLH domain in close proximity, which in the folded 'dimer' are in direct contact. On the other hand, in the 'single chain dimers' two parts of the peptides are held together that are remote from each other even in the folded conformation (Fig. 1B).

MASH-BHLH undergoes a concentration-dependent transition from a mainly unfolded monomer to a stably folded dimeric form with a dimerization constant of ~2 μM (7). On the other hand, the CD spectra of MM8 and MM17 were essentially unchanged over the concentration range 0.1–5 μM (corresponding to 0.2–10 μM monomer equivalents), as expected for a unimolecular folding reaction (Fig. 2B and data not shown). The predominant species of MM8 and MM17 are, therefore, monomers and no evidence for significant 'cross-folding' of the BHLH subunits to form dimeric species or higher aggregates or linear polymers was observed.

Structural characterization of the DNA complexes of MM8 and MM17

The sizes of the DNA complexes of MM8 and MM17 were compared with wild-type MASH-BHLH complexes in electrophoretic mobility shift assays. MCK-S, a 17 bp DNA fragment from the IgH enhancer-like element of the muscle creatine kinase gene, was used as a probe (33; Fig. 1D). Incubation of this oligonuclotide with MM8 and MM17 respectively produced mobility shifts of approximately the same magnitude as binding to dimeric wild-type MASH-BHLH (Fig. 3A), suggesting that the structures of the complexes were similar. If a single DNA binding domain had formed by cross-folding of BHLH domains from different single chain dimers retardation of the mobility of the complexes would have been significantly greater.

In order to obtain structural information the DNA complexes of MM8 and MM17 were studied by CD spectroscopy. Upon addition of 1 equiv. double-stranded oligonucleotide containing an E box sequence to a solution of MM8 or MM17 a folding transition from a largely unfolded to a mainly α-helical conformation was observed (Fig. 2A and C). A similar change in the CD spectrum occurred when MCK-S was added to wild-type MASH-BHLH (Fig. 2A; 7,8). Interestingly, the amount of helicity observed in the different complexes varied. In the DNA complex of MM17 90% of all residues were in an α-helical conformation, an increase of 5% when compared with the wild-type complex (Fig. 2A and data not shown). On the other hand, the percentage of α-helicity was ~75% in the MM8 complex (Fig. 2A). This might be a consequence of the shorter length of the linker used in MM8. Either the N-terminal end of the basic region or the C-terminal part of helix 2 might have to unfold partly to allow proper folding of MM8 on the DNA. However, if so, this local unfolding did not diminish the DNA binding affinity of MM8 (Table 1, *vide infra*).

The structural changes upon DNA binding observed in both wild-type MASH-BHLH and the 'single chain dimers' were in sharp contrast to the behaviour of disulfide-linked MASH-BHLH, which was fully folded even in the absence of DNA. No conformational change could be observed when DNA was added (7), indicating that the processes of dimerization, folding and DNA binding were uncoupled. MM8 and MM17, on the other hand, behave similarly to wild-type MASH-BHLH, in that folding and DNA binding remain coupled processes. Since the two subunits are covalently linked in the single chain dimers, no dimerization occurs on DNA binding. However, the subunits of MM8 and MM17 still undergo a conformational rearrangement which brings the two subunits into the intimate contact needed for formation of the proper complex.

Table 1. DNA binding parameters of MASH-BHLH and the 'single chain dimers' MM8 and MM17 measured by EMSA

	$[P_{1/2}]^a$ (nM)		$K_d{}^b$ (10^{15})		$\Delta G_{obs}{}^c$ (kcal/mol)		$\Delta\Delta G_{obs}{}^d$
	MCK-S[e]	SP-1[e]	MCK-S[e]	SP-1[e]	MCK-S[e]	SP-1[e]	
MASH-BHLH	458.0 (± 91)	520.0 (± 129)	209.8	270.4	–16.98	–16.84	–0.14
MM8	16.2 (± 5.5)	44.5 (± 1.1)	1.1	7.9	–10.44	–9.85	–0.59
MM17	22.3 (± 6.6)	59.3 (± 1.7)	2.0	14.1	–10.25	–9.68	–0.57

[a]Concentration of protein for which 50% of the DNA binding sites are filled. Standard deviations from multiple mesurements under identical conditions are given in parantheses.
[b]Dissociation constants are reported relative to momomer equivalents: $K_d = [\text{MASH-BHLH}]^2$ for MASH-BHLH; $K_d = (2 \times [\text{MM8}])^2$ and $K_d = (2 \times [\text{MM17}])^2$ for the 'single chain dimers.
[c]Reaction free energy for the binding reaction: $\Delta G_{obs} = -RT \ln[\text{MASH-BHLH}]^2$ for MASH-BHLH; $\Delta G_{obs} = -RT \ln[P]_{1/2}$ for MM8 and MM17. Values are for 20°C.
[d]$\Delta\Delta G_{obs} = \Delta G_{obs}(\text{MCK-S}) - \Delta G_{obs}(\text{SP-1})$ corresponds to the free energy of transfering a protein molecule from an SP-1 site to an MCK-S site.
[e]See Materials and Methods for DNA sequences.

As had previously been observed with MASH-BHLH and other BHLH proteins, the coil to α-helix transition was not only induced through addition of E box-containing DNA, but also by completely unrelated DNA (Fig. 3C; 7–9,11). Interestingly, the complex of MM8 with MCK-S contained slightly more α-helical residues than the complex with heterologous DNA. The same observation was made for the DNA complexes of MM17 (data not shown). While these observations were difficult to interpret, they nevertheless suggested a small difference in the geometry of the specific and the non-specific complexes of MM8 and MM17. It is noteworthy that no difference in the CD spectra of the specific and non-specific complexes of wild-type (8) and disulfide-linked MASH-BHLH had been observed (7).

DNA binding affinity of MM8 and MM17

Earlier work had shown that MASH-BHLH binds to DNA with moderate affinity and low DNA sequence specificity (Table 1; 7,8,11). In EMSA titration experiments, the apparent dissociation constants were measured for complexes of the 'single chain dimers' with oligonucleotides containing an E box and with completely heterologous DNA (Fig. 1D). Increasing amounts of the proteins were added to a constant amount of DNA and the extent of complex formation was measured (Fig. 3B). The protein concentration at which half of the DNA binding sites are occupied, $[P]_{1/2}$, was determined from the graphs describing the dependence of Φ, the fraction of DNA bound, on the concentration of the unbound protein (Fig. 3C).

The single chain dimers bind the MCK-S oligonucleotide half maximally at significantly lower concentrations than MASH-BHLH. While a concentration of 458 ± 91 nM MASH-BHLH was required to occupy 50% of all E boxes of MCK-S, concentrations of only 16.2 ± 5.5 and 22.3 ± 6.6 nM were needed for MM8 and MM17 respectively (Table 1). Even when the change from a monomeric to a dimeric species was taken into account, linking of the two BHLH domains lowered the half maximal binding concentration by more than one order of magnitude. The oxidized form of MASH-GGC bound to MCK-S only ~3 times tighter than wild-type MASH-BHLH (7).

The ~10-fold increase in the affinities of the single chain dimers for E box-containing DNA sequences compared with wild-type MASH-BHLH could have several different origins. The energies required for stabilization of the single chain dimers could result from additional contacts between the DNA and residues in the protein linker. While this explanation cannot be ruled out based on the existing data, the X-ray structures of the DNA complexes of E47 and MyoD (4,5) suggest that the residues of the linker pass around one side of the BHLH dimer and that they are shielded from the DNA by one of the BHLH domains (Fig. 1B). An alternative explanation is suggested by isothermal titration calorimetry experiments of the DNA binding reaction of MASH-BHLH (11). While the amino acid residues of the basic region were unfolded in the free protein even

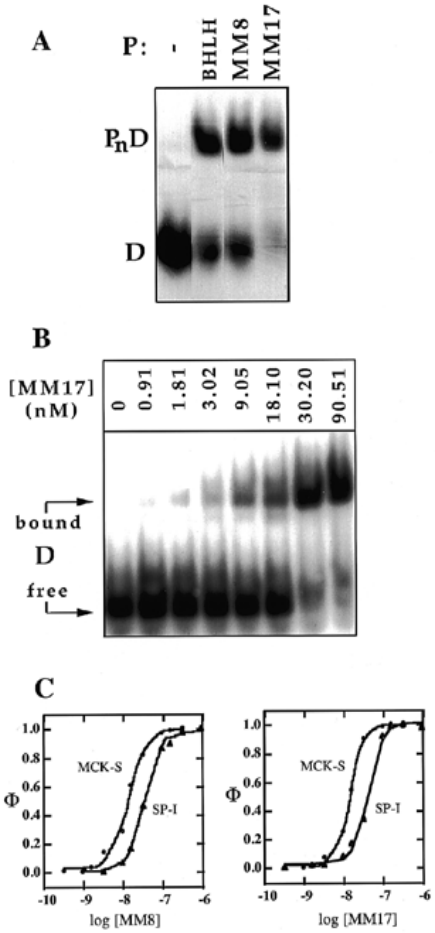

Figure 3. Characterization of the DNA complexes of MM8 and MM17 by electrophoretic mobility shift assays. (**A**) Autoradiogram of EMSAs with MCK-S oligonucleotide and various proteins. Lane 1, (–) MCK-S alone; lane 2, MCK-S with MASH-BHLH (8); lanes 3 and 4, MCK-S with MM8 and MM17 respectively. [MCK-S] 10 nM; [MASH-BHLH] 0.4 μM; [MM8] 20 nM; [MM17] 30 nM. (**B**) Autoradiogram of EMSA titration of radiolabeled MCK-S with increasing amounts of MM17. [MCK-S] 10 nM; concentrations of MM17 are indicated. (**C**) Binding isotherms for the binding of MM8 (left) and MM17 (right) to MCK-S (•) and SP-1 (▲) oligonucleotides. The symbols indicate the experimental data points for typical titrations like the one shown in (C). The lines represent the best theoretical fits through the data points. The fraction Φ of bound protein is plotted as a function of the protein concentration.

at the high protein concentrations used for ITC (where MASH-BHLH exists mainly as a stable dimer with dimerization mediated through the HLH region), the basic region adopted an α-helical conformation in the complex. In addition, NMR spectroscopy of the BHLH domain of E47 revealed that the basic region is structurally disordered in the absence of DNA (10). Therefore, the linker which connects the C-terminal end of one BHLH subunit of MM8 or MM17 to the basic region of the other might help position residues in the basic region, thereby improving contacts with the DNA. Simply by restricting the conformational freedom of the basic region in the disordered state the linker might reduce the entropic penalty that accompanies folding during DNA binding (34).

Specificity of DNA binding

Further evidence that the linker might restrict the conformational mobility of the adjacent basic region was provided by the observation that not only the DNA binding affinity but also the DNA binding specificity was increased in the single chain dimer when compared with MASH-BHLH. While the affinity for E box-containing DNA was increased in MM8 and MM17 by 10- to 14-fold, the affinity for heterologous DNA was only 4- to 6-fold higher (Table 1). As a consequence, the free energy of transferring a protein molecule from the heterologous SP-1 DNA to an oligonucleotide containing an E box was decreased from –0.14 kcal/mol for wild-type MASH-BHLH to –0.59 kcal/mol for MM8 and to –0.57 kcal/mol for MM17. Limiting the number of accessible conformations of the basic region through introduction of the linker could stabilize the complex with specific DNA to a greater extent than the complex with heterologous DNA. Interestingly, while the association reaction between the single chain dimers and MCK-S was more exergonic by ~1.2 kcal/mol than the binding reaction of disulfide-linked MASH-BHLH (Table 1; 7), the specificity increase was slightly smaller. $\Delta\Delta G_{obs}$ for MM8 and MM17 were –0.59 and –0.57 kcal/mol respectively, while for disulfide-linked MASH-BHLH $\Delta\Delta G_{obs}$ was –0.71 kcal/mol (7).

In summary, the single chain dimers MM8 and MM17 are stable, soluble, cooperatively folded proteins which bind to DNA with enhanced affinity and specificity. Unlike disulfide-linked MASH-BHLH (7), MM8 and MM17 preserve most of the characteristic DNA binding properties of wild-type MASH-BHLH. While MM8 and MM17 do not rely on dimerization for binding, they undergo substantial conformational rearrangement for DNA binding, indicating that conformational rigidity is not a requirement for enhanced DNA binding specificity of BHLH proteins.

To the best of our knowledge the linker in MM17 is the longest linker which has been used to successfully connect two protein domains (with the exception of the linkers used to create single chain antibodies). It shows that protein subunits can be successfully connected even when the appropriate C- and N-termini are remote from each other. Despite the fact that the MM17 linker must transverse >55 Å from one side of the BHLH dimer to the other, it is resistant to protease digestion in *E.coli* and does not interfere with either protein folding or DNA binding.

The single chain dimers of MASH-BHLH provide the opportunity to address several questions concerning molecular recognition. Since amino acids in the two domains can be varied independently, it should be possible through mutagenesis to direct the single chain dimers to asymmetric DNA target sequences. In addition, single chain dimers can be displayed on the surface of filamentous phage particles and new DNA binding properties can be selected for through random mutagenesis.

ACKNOWLEDGEMENTS

We thank the members of our laboratory for support and discussions, Dr Lesley A.Tannahill for critical reading of the manuscript and Dr P.L.Luisi for the use of his CD spectrometer.

REFERENCES

1 Murre,C., McCaw,P.S. and Baltimore,D. (1989) *Cell*, **56**, 777–783.
2 Davis,R.L., Cheng,P.-F., Lassar,A.B. and Weintraub,H. (1990) *Cell*, **60**, 733–746.
3 Littlewood,T. and Evan,G. (1997) *Helix–Loop–Helix Transcription Factors*. Oxford University Press, Oxford, UK.
4 Ma,P.C.M., Rould,M.A., Weintraub,H. and Pabo,C.O. (1994) *Cell*, **77**, 451–459.
5 Ellenberger,T., Fass,D., Arnaud,M. and Harrison,S.C. (1994) *Genes Dev.*, **8**, 970–980.
6 Fairman,R., Beran-Steed,R.K., Anthony-Cahill,S.J., Lear,J.D., Stafford,W.F., DeGrado,W.F., Benfield,P.A. and Brenner,S.L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 10429–10433.
7 Künne,A.G.E. and Allemann,R.K. (1997) *Biochemistry*, **36**, 1085–1091.
8 Meierhans,D., el Ariss,C., Neuenschwander,M., Sieber,M., Stackhouse,J.F. and Allemann,R.K. (1995) *Biochemistry*, **34**, 11026–11036.
9 Künne,A.G.E., Meierhans,D. and Allemann,R.K. (1996) *FEBS Lett.*, **391**, 79–83.
10 Fairman,R., Beran-Steed,R.K. and Handel,T.M. (1997) *Protein Sci.*, **6**, 175–184.
11 Künne,A.G.E., Sieber,M. and Allemann,R.K. (1998) *Biochemistry*, in press.
12 Toth,M.J. and Schimmel,P. (1986) *J. Biol. Chem.*, **261**, 6643–6646.
13 Hallewell,R.A., Laria,I., Tabrizi,A., Carlin,C., Getzoff,E.D., Tainer,J.A., Cousens,L.S. and Mullembach,G.T. (1989) *J. Biol. Chem.*, **264**, 5260–5268.
14 Bizub,D., Weber,I.T., Cameron,C.E., Leis,J.P. and Skalka,A.M. (1991) *J. Biol. Chem.*, **266**, 4951–4958.
15 Predki,P.F. and Regan,L. (1995) *Biochemistry*, **34**, 9834–9839.
16 Kim,S.-H., Kang,C.-H., Kim,R., Cho,J.M., Lee,Y.-B. and Lee,T.-K. (1989) *Protein Engng*, **2**, 571–575.
17 Tang,Y., Jiang,N., Parakh,C. and Hilvert,D. (1996) *J. Biol. Chem.*, **271**, 15682–15686.
18 Huston,J.S., Mudgett-Hunter,M.M.-S., McCartney,J., Warren,F., Haber,E. and Oppermann,H. (1996) *Methods Enzymol.*, **203**, 46–88.
19 Johnson,S. and Bird,R.T. (1996) *Methods Enzymol.*, **203**, 88–98.
20 Robinson,C.R. and Sauer,R.T. (1996) *Biochemistry*, **35**, 109–116.
21 Percipalle,P., Simoncsits,A., Zakhariev,S., Guarnaccia,C., Sanchez,R. and Pongor,S. (1995) *EMBO J.*, **14**, 3200–3205.
22 McCafferty,J., Griffiths,A.D., Winter,G. and Chiswell,D.J. (1990) *Nature*, **348**, 552–554.
23 Winter,G., Griffiths,A.D., Hawkins,R.E. and Hoogenboom,H.R. (1994) *Annu. Rev. Immunol.*, **12**, 433–455.
24 Burton,D.R. and Barbas,C.F. (1994) *Adv. Immunol.*, **57**, 191–280.
25 Rebar,E.J. and Pabo,C.O. (1994) *Science*, **263**, 671–673.
26 Jamieson,A.C., Kim,S.-H. and Wells,J.A. (1994) *Biochemistry*, **33**, 5689–5695.
27 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
28 Wetlaufer,D.R. (1962) *Adv. Protein Chem.*, **17**, 303–390.
29 Sun,X.-H. and Baltimore,D. (1991) *Cell*, **64**, 459–470.
30 Anthony-Cahill,S.J., Benfield,P.A., Fairman,R., Wasserman,Z.R., Brenner,S.L., Stafford,W.F., Altenbach,C., Hubbell,W.L. and DeGrado,W.F. (1992) *Science*, **255**, 979–983.
31 Greenfield,N. and Fasman,G.D. (1969) *Biochemistry*, **8**, 4108–4116.
32 Compton,L.A. and Johnson,W.C.,Jr. (1986) *Anal. Biochem.*, **155**, 155–167.
33 Buskin,J.N. and Hauschka,S.D. (1989) *Mol. Cell. Biol.*, **9**, 2627–2640.
34 Spolar,R.S. and Record,M.T. (1994) *Science*, **263**, 777–784.
35 Davis,R.L., Weintraub,H. and Lassar,A.B. (1987) *Cell*, **51**, 987–1000.
36 German,M.S., Blanar,M.A., Nelson,C., Moss,L.G. and Rutter,W.J. (1991) *Mol. Endocrinol.*, **5**, 292–299.
37 Bacon,D.J. and Anderson,W.F. (1988) *J. Mol. Graphics*, **6**, 219–220.
38 Meritt,E.A. and Murphy,M.E.P. (1994) *Acta Crystallogr. D*, **50**, 869–873.