Chemistry & Biology, Vol. 11, 69–77, January, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.12.015

Controlling the DNA Binding Specificity of bHLH Proteins through Intramolecular Interactions

Elizebeth C. Turner,[1] Charlotte H. Cureton,[2] Chris J. Weston,[2] Oliver S. Smart,[2] and Rudolf K. Allemann[1,*]
[1]School of Chemistry
[2]School of Biosciences
University of Birmingham
Edgbaston
Birmingham, B15 2TT
United Kingdom

Summary

Reversible control of the conformation of proteins was employed to probe the relationship between flexibility and specificity of the basic helix-loop-helix protein MyoD. A fusion protein (apaMyoD) was designed where the basic DNA binding helix of MyoD was stablized by an amino-terminal extension with a sequence derived from the bee venom peptide apamin. The disulfide-stabilized helix from apamin served as a nucleus for a helix that extended for a further ten residues, thereby holding apaMyoD's DNA recognition helix in a predominantly α-helical conformation. The thermal stability of the DNA complexes of apaMyoD was increased by 13°C relative to MyoD-bHLH. Measurements of the fluorescence anisotropy change on DNA binding indicated that apaMyoD bound to E-box-containing DNA sequences with enhanced affinity relative to MyoD-bHLH. Consequently, the DNA binding specificity of apaMyoD was increased 10-fold.

Introduction

DNA binding proteins have evolved to deal with a problem not normally encountered by enzymes: namely that the substrate, which is a specific DNA fragment, is immersed in a sea of other DNA sequences on the same molecule that are chemically and structurally very similar to the specific substrate [1]. This sequence discrimination is often the basis of significant physiological differences. The production of the basic helix-loop-helix (bHLH) transcription factor MyoD in a wide variety of cell types, including fibroblasts and myoblasts, activates a cascade of genes eventually leading to cellular differentiation and the production of muscle cells [2, 3], while production of the bHLH protein MASH-1 promotes differentiation of committed neuronal precursor cells [4]. The physiological activities of MyoD and MASH-1 depend on the presence of DNA sequences containing the symmetrical core motif CANNTG (E box) [5], to which bHLH proteins bind as homodimers (Figure 1B) or heterodimers with the E12 and E47 proteins [6, 7, 8].

In stark contrast to their high physiological specificities, MASH-1 and MyoD display only limited DNA binding specificity in vitro [9, 10]. Electrophoretic mobility shift assays and isothermal titration calorimetry revealed that the apparent dissociation constants of complexes with E-box-containing DNA sequences and of complexes with completely unrelated DNA sequences were similar [9, 11, 12]. Hence, the specificity of transcriptional activation needed to explain the exquisite physiological specificity of MASH-1 and MyoD cannot be based solely on their intrinsic DNA binding specificities but might be achieved through cooperative interactions with other components of the transcriptional machinery such as myocyte enhancer factors-2 (MEF-2) [13]. While not intrinsically myo- or neurogenic, the proteins of the MEF-2 family have been shown to act as coregulators to potentiate the myogenic and neurogenic effects of MyoD and MASH-1 [14, 15].

The DNA binding reactions of bHLH proteins are characterized by a transition from a largely unfolded to a mainly α-helical conformation of the bHLH domain [9, 11]. The limited DNA binding specificity of bHLH proteins has been postulated to be a consequence of their conformational flexibility due to high solvent accessibility [16]. Limiting the number of accessible conformations of the residues in the basic region of bHLH proteins, for instance, through the interaction with other proteins, should increase their DNA binding specificity by stabilizing the interaction with DNA sequences containing the preferred binding site or destabilizing the complex with heterologous DNA sequences. Herein we report the construction of a model system for the modulation of the DNA binding specificity of bHLH proteins through intramolecular interactions in which the DNA recognition helix of MyoD was stabilized through an additional N-terminal disulfide-stabilized motif (Figure 1).

Results

Design of ApaMyoD

To test the hypothesis that increasing the stability of the DNA recognition helix of MyoD enhances its DNA binding specificity, a hybrid protein, apaMyoD, was designed in which a sequence based on the peptide apamin was fused to the N terminus of the bHLH domain of MyoD. Apamin is an 18-residue neurotoxic peptide from honeybee venom in which two turns of the C-terminal α helix are stabilized through two disulfide bonds. These confer significant conformational stability to the peptide [17, 18, 19] and make this peptide an ideal template both for helix stabilization [20] and the display of functional epitopes [21].

In this study, the apamin fold was used to stabilize the basic recognition helix at the N terminus of the bHLH domain of MyoD. Molecular modeling was undertaken to carefully design the hybrid. This modeling took as a starting point the X-ray structure of the bHLH domain of MyoD complexed to DNA [22] (Figure 1B) and a model based on a recent NMR study of apamin (Figure 1C) [19, 23, 24]. Previous studies had indicated that the eight N-terminal amino acids and cysteines 11 and 15 were

*Correspondence: r.k.allemann@bham.ac.uk

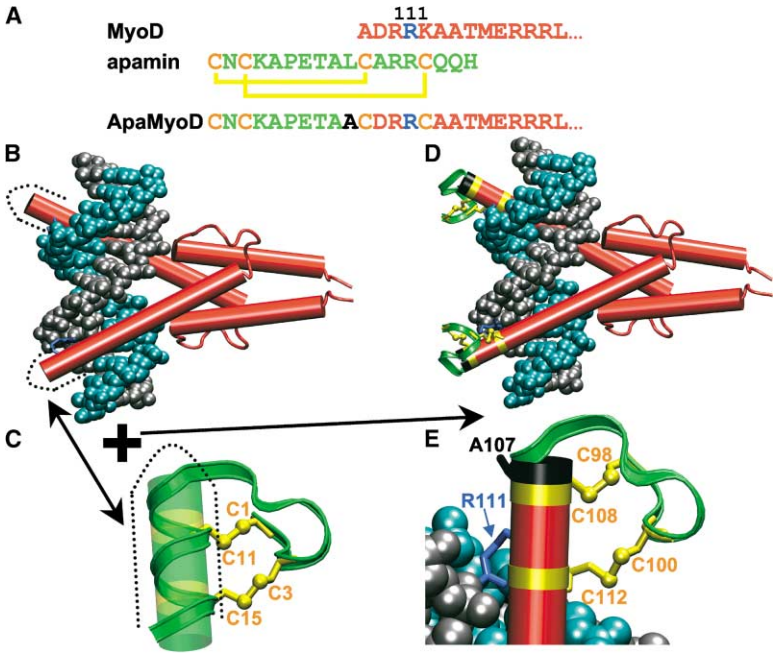

Figure 1. Design of ApaMyoD from Apamin and MyoD-bHLH

(A) Sequence alignment of the bHLH domain from MyoD, apamin, and the resultant fusion protein apaMyoD.

(B) Structure of the bHLH domain of MyoD bound as a dimer (red) to cognate DNA (gray/blue van der Waals spheres) [22]. α helices are represented by cylinders, with tubes marking nonhelical parts of the structure.

(C) Structure of the bee venom peptide apamin from NMR studies [20, 23, 24]. Atoms forming the disulfide bonds that stabilize the structure are shown in a yellow stick representation, with the rest of the structure marked as a green ribbon.

(D and E) The expected architecture of apaMyoD when bound to DNA. The parts of the hybrid derived from MyoD are marked in red or blue, whereas apamin-derived peptide segments are marked in green or yellow. The single residue not drawn from a parent sequence, A107, is marked in black. The side chain of the important nucleobase-contacting residue R111 is shown as a blue stick model. A close-up view of the end of a single chain of the hybrid (E) shows that the apamin extension points away from the DNA while holding the conformation of the N-terminal part of the basic DNA recognition helix.

necessary for the correct folding of apamin-based hybrids, while the other residues within the α-helical section at the C terminus could be varied widely [20]. Accordingly, a number of possible hybrids of the two structures were considered by using least squares fitting of the α-helical parts from each structure with the Sybyl 6.7 package. An important consideration was that no change should be made to any amino acid residue from MyoD that had been shown to directly contact DNA in the X-ray complex structure [22]. Furthermore, hybrids in which the apamin extension approached the DNA duplex were also precluded. These considerations resulted in the hybrid apaMyoD, whose sequence is shown in Figure 1A. A molecular model of the hybrid showed that the apamin fold was placed over the end of the basic helix from MyoD (Figures 1D and 1E). To ensure that the stabilizing effect of the apamin extension was in close proximity to the DNA binding region of MyoD, the solvent-exposed residue K112 was replaced by cysteine and 11 amino acids of apamin were spliced onto the basic domain, N-terminal to residue 109 (Figure 1A). The X-ray structure of a DNA complex of MyoD [22] indicated that residue R111 made a nucleobase-specific contact. The apamin extension was placed around this position so it could serve chiefly to stabilize the basic helix of apaMyoD.

Only one amino acid of the hybrid did not derive from either the sequence of MyoD or apamin: residue 107, which would naturally be a leucine (number 10) in apamin (Figure 1A). The molecular model of apaMyoD (Figure 1E) indicated that this position would be solvent exposed in the hybrid and would be positioned close to the important DNA base contacting residue R111 [22]. This residue was therefore changed to alanine both to remove an exposed hydrophobic residue and to prevent any interference of the side chain with R111. It can be noted that the hybrid apamin-based peptides with an alanine at this position were found to have three-dimensional structures essentially identical to native apamin [20]. Furthermore, it is known that mutation of leucine 10 in apamin to alanine does not effect the neurotoxicity [25] of the peptide, once again indicating that the structure is not affected by such a change.

The design process and the molecular model for ApaMyoD in complex with its cognate DNA (Figure 1B) indicated that the hybrid protein should adopt a similar DNA binding conformation to the natural protein. What can be expected for the conformation of ApaMyoD prior to DNA binding? As noted above, bHLH proteins are found largely in unfolded conformations in the absence of DNA [9, 11]. It can be expected that the $\sim$8 residues found to have a helical conformation in apamin (TAxCxxxC) [19] would also be in a helical conformation in the hybrid. Furthermore, as helix formation is cooperative [26], apamin provides an excellent nucleation point for further helix stabilization. In order to gauge the magnitude of the effect that could be expected, we undertook calculations with the AGADIR program for the prediction of the helicity of isolated peptides [27] (inspired by similar considerations made by Wemmer and coworkers [20]). The calculations (data not shown) indicated that prior to DNA binding the apamin extension has the potential to stabilize up to a further two turns of helix in addition to those directly contributed by apamin.

Synthesis and Structural Characterization of ApaMyoD

ApaMyoD was produced in *E. coli* and purified according to a procedure established previously for the bHLH domains of MASH-1 and MyoD [9, 10]. ApaMyoD

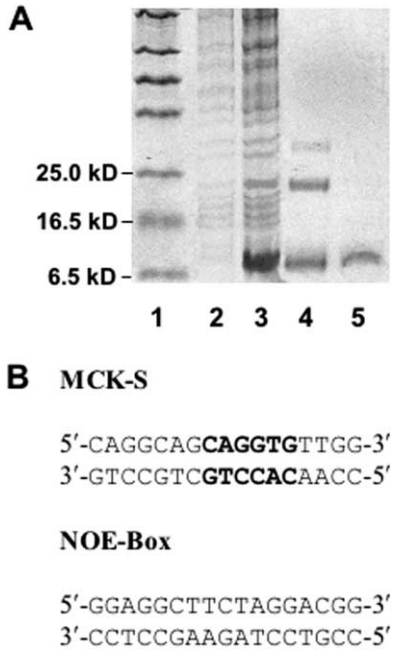

Figure 2. Expression and Purification of the Hybrid ApaMyoD

(A) Expression and purification of apaMyoD visualized by SDS-PAGE. Lane 1, molecular weight markers; lane 2, protein extract from BL21(DE3) cells containing the apaMyoD expression plasmid before induction of expression by IPTG; lane 3, protein extract 3 hr after induction; lane 4, pooled apaMyoD containing fractions after CM chromatography; lane 5, purified apaMyoD after FPL chromatography on Resource-S cation exchange resin.

(B) Base sequences of double-stranded oligonucleotides containing cognate (MCK-S) and noncognate (NOE box) binding sites.

was isolated successfully by cation exchange chromatography through elution with a shallow gradient of NaCl and was essentially pure as judged by SDS-polyacrylamide gel electrophoresis (Figure 2A). Mass spectrometry and titration with Ellman's reagent indicated that purified apaMyoD was in its oxidized, monomeric form (data not shown).

CD spectroscopy had revealed previously that the bHLH domain of MyoD was largely unfolded at concentrations below 5 μM [9]. The mean residue ellipticity at 222 nm, $[\Theta]_{r,222}$ was determined as $-3.25\ (\pm 0.3) \times 10^3$ deg cm^2 dmol^{-1} for MyoD-bHLH (1 μM), which corresponded to an α-helical content of approximately 7% or four amino acid residues (Figure 3A). At the same concentration, the CD spectrum of apaMyoD showed minima at 208 nm and 222 nm, typical of proteins with α-helical content (Figure 3A) [28]. The value of $-9.17\ (\pm 0.6) \times 10^3$ deg cm^2 dmol^{-1} measured for $[\Theta]_{r,222}$ indicated that approximately 27% of apaMyoD or 19 amino acid residues were in an α-helical conformation. This value was slightly higher than would have been expected from the addition of a disulfide-stabilized apamin-like segment containing approximately two α-helical turns. The apamin helix most likely served as a nucleus for a helix that extended for a further ten residues; the whole of the basic DNA recognition helix was apparently held in a predominately helical formation by the apamin extension.

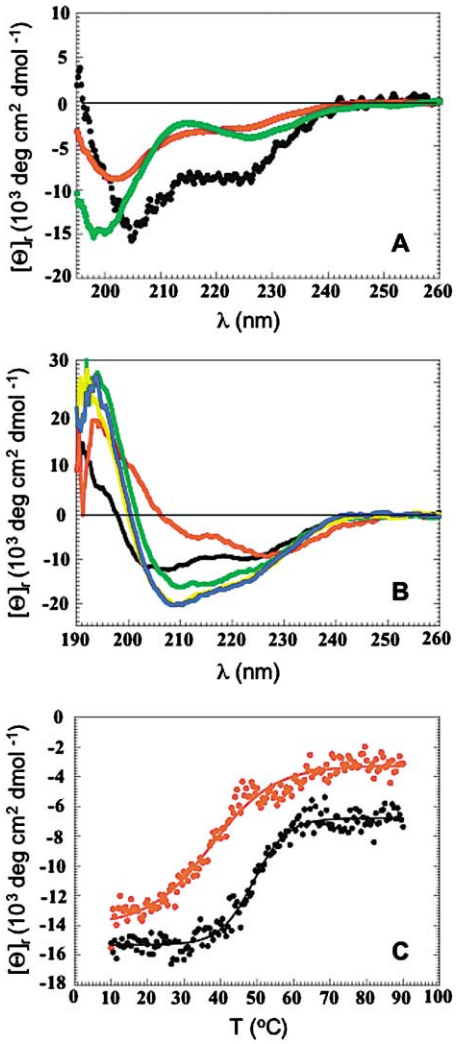

Figure 3. Structural Characterization of ApaMyoD by CD Spectroscopy

(A) CD spectra of MyoD-bHLH (red), oxidized apaMyoD (black), and reduced apaMyoD (green) in the absence of DNA. Protein concentrations were 1 μM for all experiments.

(B) Effects of increasing concentrations of cognate DNA (MCK-S) on the CD spectrum of apaMyoD. CD spectra of apaMyoD (1 μM) with 0 μM (black), with 0.1 μM (red), with 0.3 μM MCK-S (green), with 0.5 μM (yellow), and with 0.8 μM MCK-S (blue) are shown.

(C) Temperature dependence of the mean residue ellipticity, $[\Theta]_r$, at 222 nm of 4 μM apaMyoD (black) and MyoD-bHLH (red) bound to cognate DNA ([MCK-S] = 2 μM).

ApaMyoD could be fully reduced with tris(2-carboxyethyl) phosphine (TCEP). This was demonstrated with Ellman's reagent. The CD spectrum of the reduced form of apaMyoD was similar to that of its parent protein MyoD-bHLH (Figure 3A), indicating that the stabilization of the secondary structure of apaMyoD was a direct result of the formation of the disulfide bridges.

Structural Characterization of the DNA Complexes of ApaMyoD by CD Spectroscopy

CD spectroscopy was also used to analyze the formation of complexes between apaMyoD and DNA. The

addition of the MCK-S oligonucleotide comprising 17 base pairs of the IgH enhancer-like element of the muscle-specific creatine kinase enhancer [29] with the central E-box sequence CAGGTG (Figure 2B) to apaMyoD (1 μM) in its oxidized form induced a change in the CD spectra indicative of a transition from a partially unfolded to a mainly α-helical conformation. A similar transition had been observed previously with the bHLH domain of MyoD [10] and with MASH-bHLH [9]. The change in the CD spectrum was saturable, in that the addition of excess oligonucleotide (0.8 μM MCK-S and 1 μM apaMyoD monomer, equivalent to 0.5 μM dimer) did not result in a further change of the spectrum relative to a 1:1 stoichiometry (0.5 μM apaMyoD dimer: 0.5 μM DNA duplex). The transition of the conformation of apaMyoD was not dependent on the DNA sequence, as demonstrated in experiments with the noncognate oligonucleotide, NOE box (data not shown).

Heat denaturation of the DNA complexes of oxidized apaMyoD was monitored through changes in the CD signal at 222 nm. The DNA complexes of oxidized apaMyoD with both MCK-S and the NOE box oligonucleotide underwent a cooperative unfolding reaction with a midpoint of 50°C, which is 13°C higher than the melting temperature observed for the DNA complexes of the bHLH domain of MyoD (Figure 3C), indicating that the addition of the apamin extension conferred significant additional stability to the DNA complexes of MyoD.

DNA Binding Specificity Measured by Fluorescence Anisotropy

The DNA binding properties of the bHLH domain of MyoD had been studied previously by electrophoretic mobility shift assays (EMSA), which revealed an almost complete lack of sequence specificity [10]. The thermodynamically rigorous technique of fluorescence anisotropy [30–32] is premised on the differences in size, and hence the differences in the rates of rotational and fluorescence anisotropies, of the free oligonucleotide labeled with a fluorescent tag and its protein complex. It therefore allows the determination of the apparent dissociation constants of protein-DNA complexes under true equilibrium conditions. Typical results from such fluorescence polarization experiments are shown in Figure 4. The addition of protein to a solution of fluorescently labeled double-stranded oligonucleotide resulted in a saturable increase in the fluorescence anisotropy, indicating that the protein was bound to the DNA in solution. In agreement with the results from EMSA experiments [10], MyoD-bHLH showed similar affinities for MCK-S and the heterologous DNA sequence NOE box (Figure 4A). The NOE-box oligonucleotide has the same base composition as MCK-S but with a scrambled sequence, and it does not contain an E-box sequence. The bHLH proteins MASH-1 and MyoD have been shown previously to bind to NOE-box DNA with significantly reduced affinity relative to MCK-S [9–12]. The K_D values for MyoD-bHLH binding to MCK-S and NOE box were determined as 8.0 (±0.4) × 10^{-16} M^2 and 9.0 (±0.8) × 10^{-16} M^2, respectively, resulting in an energetic difference at 25°C between the stability of the two complexes of only 70 cal mol^{-1} (Table 1).

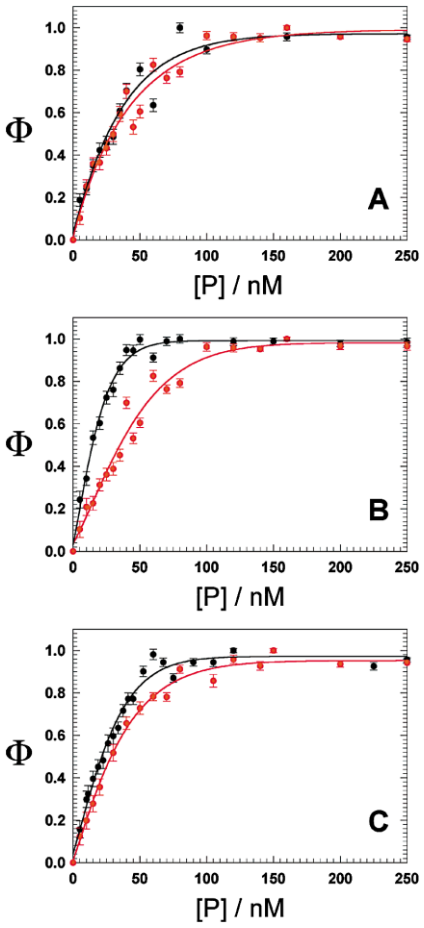

Figure 4. Characterization of DNA Binding by ApaMyoD Measured by Fluorescence Anisotropy

Fraction, ϕ, of bound DNA in the complexes of (A) MyoD-bHLH, (B) oxidized apaMyoD, and (C) reduced apaMyoD versus the concentration of the free protein. The binding curves for complexes with 1 nM solutions of cognate (MCK-S) and noncognate DNA (NOE box) are indicated in black and red, respectively.

The specificity of DNA binding by apaMyoD was significantly increased compared to that of MyoD-bHLH. The concentration of oxidized apaMyoD needed to bind 50% of MCK-S was approximately four times smaller than that of MyoD-bHLH with a dissociation constant of 4.1 (±0.5) × 10^{-17} M^2 (Table 1). The presence of the apamin extension clearly increased the affinity of apaMyoD for specific, E-box-containing DNA. Control experiments with apamin exhibited no change in fluorescence anisotropy over the whole range of peptide concentrations studied with MyoD-bHLH and apaMyoD, indicating that apamin itself had DNA binding affinity significantly lower than apaMyoD. The increased DNA binding affinity of oxidized apaMyoD was dependent on the presence of the disulfides; when the disulfides were reduced with TCEP, the dissociation constant of the complex of MCK-S with reduced apaMyoD was 4.0 (±0.2) × 10^{-16} M^2 (Table 1). Interestingly, even in its reduced form apaMyoD bound cognate DNA slightly better than MyoD-bHLH, possibly due to residual structure of the apamin extension even in the absence of

Table 1. DNA Binding Parameters for MyoD-bHLH and Oxidized and Reduced ApaMyoD Determined by Fluorescence Anisotropy

	MCK-S[a]			NOE box[a]				
	$[P]_{i/2}$[b] (nM)	K_D[c] (M^2)	ΔG^d (kcal mol^{-1})	$[P]_{i/2}$[b] (nM)	K_D[c] (M^2)	ΔG^d (kcal mol^{-1})	$\Delta\Delta G^e$ (kcal mol^{-1})	Specificity[f]
MyoD-bHLH	28.3 (±0.7)	$8.0\ (\pm0.4) \times 10^{-16}$	20.57 (±0.03)	30.0 (±1.3)	$9.0\ (\pm0.8) \times 10^{-16}$	20.50 (±0.05)	−0.07	1.13
$ApaMyoD_{Red}$	20.0 (±0.4)	$4.0\ (\pm0.2) \times 10^{-16}$	20.98 (±0.03)	27.7 (±0.8)	$7.7\ (\pm0.5) \times 10^{-16}$	20.59 (±0.04)	−0.39	1.93
$ApaMyoD_{Ox}$	6.4 (±0.4)	$4.1\ (\pm0.5) \times 10^{-17}$	22.33 (±0.07)	21.2 (±0.9)	$4.5\ (\pm0.4) \times 10^{-16}$	20.91 (±0.05)	−1.42	10.98

[a] See Figure 2B for DNA sequences.
[b] Protein concentration for which 50% of the DNA binding sites are filled.
[c] Apparent dissociation constants are reported as $K_D = ([P]_{i/2})^2$. They are the average of at least three independent measurements, and the standard errors are given in parentheses.
[d] $\Delta G = -RT \ln K_D$.
[e] $\Delta\Delta G = \Delta G$ (MCK-S) − ΔG (NOE box).
[f] Specificity of the DNA binding reaction is defined as the ratio K_D(NOE box)/K_D(MCK-S).

the disulfide bonds [24]. The small increase in the DNA binding affinity of the reduced form of apaMyoD may have resulted from the short extension of the basic helix at its N terminus through the introduction of the apamin residues.

The affinity of oxidized apaMyoD for the noncognate NOE box DNA was increased to a much smaller extent by the apamin extension [$K_D = 4.5\ (\pm0.4) \times 10^{-16}\ M^2$] than had been observed for MCK-S (Table 1). Overall, the introduction of the disulfide linkage between apamin and the basic region of MyoD-bHLH led to a significantly increased stability of the complexes of apaMyoD with specific DNA and an increase in the DNA binding specificity of one order of magnitude. The specificity displayed by reduced apaMyoD was increased less than 2-fold relative to MyoD-bHLH (Table 1).

Discussion

The DNA binding regions of eukaryotic transcription factors are often not part of a globular fold, but isolated elements of secondary structure. The members of the basic-zipper (BZ), the bHLH, and the bHLH-zipper families all interact with the major groove of DNA through an α helix [33, 34], large parts of which are solvent exposed even in the complexes. These recognition helices adopt well-defined structures only upon binding to DNA; in their unliganded forms they show significant conformational flexibility. Proteins that rely on BZ, bHLH, or bHLH-zipper elements for DNA binding normally display only limited sequence specificity in vitro. The DNA binding specificities of prokaryotic transcriptional regulators, on the other hand, are generally high. Their recognition elements are normally part of a compact DNA binding domain. The recognition helix of the prokaryotic helix-turn-helix (HTH) motif, for instance, is stabilized through interactions with other parts of the protein [35, 36]; indeed, the HTH motif adopts a stable conformation only in the structural context of the whole protein [37, 38]. A correlation between the extent of DNA binding specificity of a protein and the conformational rigidity of its DNA recognition element has been proposed previously [39]. Shepartz and coworkers grafted residues from the DNA recognition helix of the *engrailed* homeodomain onto the helix of avian pancreatic polypeptide [40]. The achieved structural preorganization compensated thereby for the reduction of the free energy of binding as a result of the reduced number of protein-DNA contacts relative to the DNA complex of the parent homeodomain.

The low DNA binding specificities of many eukaryotic transcription factors are in sharp contrast to their exquisite physiological specificities. The bHLH protein MyoD, for instance, has been called the master regulator of muscular differentiation because its expression induces the differentiation of many cell types into myotubes [41, 42]. Transcriptional regulation in eukaryotes relies largely on multiprotein complexes with the potential for combinatorial interactions that can change the conformational flexibilities of DNA recognition elements, leading to variations in the DNA binding specificities.

As a model for such interactions of the recognition

helix of MyoD with other components of the transcriptional machinery, we have produced apaMyoD, in which a N-terminal extension to the basic region of MyoD stabilized the α helix through two disulfide bonds. ApaMyoD bound to E-box-containing DNA sequences more avidly than its parent protein MyoD-bHLH (Table 1). bHLH proteins are known to undergo a conformational change on DNA binding from a largely unfolded to a mainly α-helical structure [9]. The basic region is unfolded even at concentrations where the protein is mainly dimeric and the HLH-region stably folded [12]. On DNA binding, the conformation of these residues becomes helical too, and previous work for the DNA binding reaction of MASH-1 indicated that a significant entropic penalty resulted from this folding transition [12]. CD spectroscopy had indicated that approximately 27% of apaMyoD was in an α-helical conformation at DNA binding concentrations, while only 7% of those of MyoD-bHLH were helical. The increased DNA binding affinity of apaMyoD relative to MyoD-bHLH was most likely the result of the increased proportion of α-helical structure within the recognition helix, leading to a reduction of the number of residues undergoing a conformational change during DNA binding.

The stability of the DNA complexes of apaMyoD to thermal denaturation was also significantly increased. For MyoD-bHLH, an unfolding transition was observed at 37°C in thermal denaturation experiments. For the DNA complexes of apaMyoD, on the other hand, this transition occurred at 50°C. The unfolding of apaMyoD above this temperature was most likely due to the lack of stability of the complex rather than a consequence of the thermally induced melting of the oligonucleotide, which occurs with a midpoint of 61°C [12].

Interestingly, the affinity of apaMyoD for heterologous DNA sequences was increased only approximately 2-fold. Limiting the number of accessible conformations of the recognition helix of apaMyoD appeared to stabilize the interaction with specific DNA and destabilize the complex with nonspecific DNA, leading to a significant increase in the DNA binding specificity of the hybrid protein. This enhanced specificity was clearly a consequence of the disulfide-stabilized intramolecular interactions, as the reduced form of apaMyoD displayed specificity similar to that of MyoD-bHLH (Table 1). Local stabilization of the DNA binding domain the tryptophan repressor of *E. coli* [43] has been shown to lead to a reduction of the conformational flexibility and increased specificity. Amino acid substitutions that stabilized the DNA binding domain of the repressor locally led to a reduction of the conformational flexibility [44] and an increase in the DNA binding specificity through reduced affinity for nonspecific DNA [45].

The intramolecular interactions between the N-terminal apamin-like extension and the basic region in apaMyoD can be seen as a model for the modulation of the DNA binding properties of transcriptional regulators through intermolecular interaction between their DNA recognition elements and other components of the transcription complex such as the members of the MEF-2 family. MEF-2C, a potent coregulator of MyoD, could stabilize a specific conformation of the basic region of MyoD, thereby altering its intrinsic DNA binding specificity. Such a mechanism could help to resolve the discrepancy between the low DNA binding specificity of MyoD in vitro and its exquisite physiological specificity. A similar mechanism has been proposed for the activation by human T cell leukaemia virus type I Tax protein of many cellular transcription factors that depend on a BZ domain for DNA recognition [46, 47]. Tax alters the relative affinity of basic-zipper proteins for different DNA binding sites through direct interaction with their basic domains, thereby increasing the stability of the dimeric forms of BZ proteins [48, 49]. Similarly, the hepatitis B virus protein pX has been shown to increase the stability of the DNA complexes of BZ proteins in a sequence-dependent manner [50]. Thus, Tax and pX increase the DNA binding affinity and specificity of BZ proteins through a mechanism similar to that of apaMyoD.

The work described in this paper represents the first example of a designed transcription factor that binds to DNA with significantly increased DNA binding affinity and specificity as a consequence of disulfide-dependent intramolecular stabilization of the DNA recognition helix. Our results provide a possible resolution of the mechanistic paradox between the impressive physiological specificity displayed by bHLH proteins and their modest DNA binding specificity in vitro.

Significance

In sharp contrast to many prokaryotic transcription factors, eukaryotic transcriptional regulators often display only limited sequence specificity in vitro due to their relatively simple structures and the high conformational flexibilities of their DNA binding motifs. The significant stabilization of the DNA recognition helix of the basic helix-loop-helix protein MyoD that resulted from the intramolecular interaction with an N-terminally fused apamin-like peptide led to a redox-state-dependent increase in the affinity of the bHLH motif for its natural DNA target sites. As a consequence, the specificity of DNA binding also increased. The fusion protein between MyoD and apamin provided a model for the modulation of the DNA binding properties of proteins like MyoD by intermolecular interaction with other members of the transcription complex. Because the stabilization of the basic region of MyoD is achieved mainly through the formation of two disulfide bonds, the properties of the fusion protein could be controlled through changes in the redox conditions. Many other biologically important processes that rely on the interaction with key helical domains might be reversibly controlled in a similar fashion.

Experimental Procedures

Materials

All reagents were from Sigma and Invitrogen unless indicated otherwise. Oligonucleotides were purchased from Alta-Biosciences, University of Birmingham. Sequences of oligonucleotides used are shown in Figure 2B. Complementary single-stranded oligonucleotides were annealed by heat denaturation followed by slow cooling to room temperature. Oligonucleotides used for fluorescence measurements were 3′ labeled with acetamido-5-fluorescein and subse-

quently annealed to their complementary unlabeled oligonucleotide as described previously [12].

Molecular Modeling

Modeling was performed using the Sybyl 6.7 program (Tripos) running on a Silicon Graphics Octane workstation. The 2.8 Å X-ray structure of the bHLH domain of MyoD complexed to cognate DNA [22] was obtained from the Protein Data Bank (ID code 1mdy). Coordinates for apamin were based on a recent NMR structure [23]. This was adjusted to incorporate likely interactions for Thr 9 [19] and Asn 2 [24]. Furthermore, the conformations of the last three residues that tend to be disordered in apamin were set to ideal α-helical.

Molecular graphics visualization of the two parent structures was used to decide upon the sequence of the hybrid protein. The two structures were superimposed using least squares fitting to overlay the main chain atoms of α-helical parts. Once a sequence had been chosen (Figure 1A), a model for the hybrid protein was constructed (Figure 1). This involved using least squares fitting to place the main chain of residues A12 to H18 from apamin over corresponding atoms from MyoD residues D109 to T115. The extension to the structure drawn from apamin was then merged with the MyoD molecule, and residues 108 and 112 were altered to cystine within the Sybyl program. The modified residues were then subjected to local energy minimization to regularize bond lengths and angles using the conjugate gradients protocol and an AMBER potential energy function [51] within Sybyl. The resultant model is conceptual, demonstrating that the apamin extension of the hybrid can be accommodated without disrupting any DNA-recognition interactions found for the native bHLH domain of MyoD (Figure 1E). Molecular graphics images were produced using the VMD program [52].

Expression of MyoD-bHLH and ApaMyoD

A cDNA for apaMyoD was constructed in the plasmid pJGetMyoD-bHLH that had been used previously for the production of the bHLH domain of MyoD [10]. pJGetMyoD-bHLH contains the cDNA for the bHLH domain of MyoD within plasmid pJGetita [10], a derivative of pET3a [53], thereby allowing high-level expression from the T7 promoter in BL21(DE3) cells. The 5′ extension coding for the apamin addition (Figure 1) was inserted upstream of the basic region of MyoD-bHLH by PCR using the following primers: A (5′-GTCCATAT GTGCAACTGCAAAGCTCCCGAGACCGCTGCTTGCGCTGATCGCC GCAAGGCC-3′) and B (5′-AGCGGATCCTCATCAGTCGCGCAGC AGCTGCAG-3′). Following digestion with NdeI and BamHI, the PCR product was inserted into pJGetita. The codon for Lys112 was subsequently replaced with the codon for Cys by PCR using the following primers: C, (5′-GATATACATATGTGCAACTGCAAAGCTCCC-3′); D, (5′-CCTGCGCTGATCGCCGCTGCGCCGCCACCATG-3′); E, (5′-CATGGTGGCGGCGCAGCGGCGATCAGCGCAAG-3′); and F, (5′-AGC GGATCCTCATCAGTCGCGCAGCAGCTGCAG-3′). Altered nucleotides are underlined. The DNA sequence was verified using the dideoxy sequencing method [54]. BL21(DE3) cells containing the apaMyoD expression plasmid were grown at 37°C in LB media containing 100 μg ml^{-1} ampicillin until the OD_{600} reached 0.5. IPTG was added to a final concentration of 1 mM, and the cells were harvested after 3 hr by centrifugation.

Protein Purification

MyoD-bHLH was purified as described previously [10]. ApaMyoD was purified essentially as described for the purification of the bHLH domain of MASH-1 [9]. In short, the cells were resuspended in 3 ml of water/g of wet cells, and 1 mM phenylmethanesulfonyl fluoride was added. Following the addition of 2 volumes of lysis buffer (100 mM ammonium acetate [pH 6.7], 100 mM sodium chloride, 100 mM 2-mercaptoethanol), the cells were sonicated for 10 min on ice. The resulting suspension was dialyzed at ambient temperature twice against urea buffer (5 mM sodium acetate [pH 5.0], 100 mM 2-mercaptoethanol, 8 M urea). The dialysate was centrifuged and the supernatant was applied to a column containing 30 ml of Bio-Gel CM ion-exchange resin (Bio Rad) preequilibrated with urea buffer. The loaded resin was washed extensively with urea buffer, and the protein was eluted with one column volume of urea buffer containing 1 M sodium chloride. The eluate was dialyzed twice against urea buffer. The protein was further purified by preparative FPLC on a Resource-S (Pharmacia) ion-exchange column as described [9]. The collected fractions containing apaMyoD were pooled and concentrated by ultrafiltration using an Amicon YM-3 filter. The buffer was exchanged to 5 mM sodium acetate (pH 5.0) by dialysis. SDS-PAGE showed a single protein band (Figure 2A). Titration with Ellman's reagent showed that apaMyoD had oxidized during the final dialysis step [55]. MALDI-TOF mass spectrometry revealed a mass of 8052.0, which corresponded well with the calculated mass of 8050.5 for the oxidized monomer of apaMyoD without the N-terminal Met. Protein concentrations were determined using the microtannin assay [56]. The yields for the preparation were approximately 0.5 mg of purified protein/l culture. Reduced apaMyoD was prepared from the oxidized form by reduction with TCEP. A solution of apaMyoD (5 μM) was treated with TCEP (2.5 mM) to give a fully reduced sample as judged by CD and Ellman's reagent.

CD Spectroscopy

Spectra were all measured at 20°C (unless otherwise indicated) using a Jasco J810 spectropolarimeter. The concentrations of MyoD-bHLH and apaMyoD were 1 μM in 5 mM Tris-HCl (pH 7). For thermal denaturation studies, the protein concentration was increased to 4 μM, and the sample was heated from 5°C to 90°C at a rate of 0.5°C min^{-1}. The CD results are reported as mean residue ellipticities $[\Theta]_r$, which was calculated from the observed ellipticity, Θ, measured in mdeg according to the following equation:

$$[\Theta]_r = \Theta/10\ (n\text{-}1)\ c\ l,$$

where c is the concentration of the peptide in mol l^{-1}, n is the number of amino acids in the peptide, and l is the path length in cm. Values reported for mean residue ellipticities are the average of at least three independent measurements.

Fluorescence Anisotropy

Fluorescence anisotropy measurements were made at 25°C on a Perkin Elmer Luminescence Spectrometer LS50B arranged in L format (494 nm excitation, 525 nm emission). Titrations were performed in 0.5 ml quartz cuvettes. The assay buffer was 5 mM Tris-HCl (pH 7.9), 150 mM NaCl, 6 mM $MgCl_2$, and 10% glycerol. Defined volumes (0.5–2 μl) from a stock solution of the protein (2.5-10 μM) were added successively to 1 nM solutions of the fluorescently labeled double-stranded oligonucleotide (3′ labeled with acetamido-5-fluorescein) in a total volume of 1 ml. The G factor (ratio of sensitivities of the monochromator for horizontally and vertically polarized light) was calculated for each measurement using the following equation [57]:

$$G = I_{\perp}/I_{||}, \quad (1)$$

where $I_{||}$ and $I_{\perp}$ are the intensities of the fluorescent emissions in parallel and perpendicular planes, respectively, to the excitation plane. The G factor values ranged from 0.99–1.06. Values for fluorescence anisotropy (A) were then determined from the following equation [58]:

$$A = (I_{||} - GI_{\perp})/(I_{||} + 2GI_{\perp}). \quad (2)$$

For each anisotropy value, ten measurements were taken and averaged using an integration time of 5 s.

Fluorescence data were expressed (see Figure 4) as

$$\phi = (A\text{-}A_D)/(A_{PD}\text{-}A_D), \quad (3)$$

where A denotes the fluorescence anisotropy in the presence of the indicated concentration of protein, A_D denotes the fluorescence anisotropy of the solution containing the fluorescently labeled DNA and no protein, and A_{PD} denotes the fluorescence anisotropy at saturation [32, 58]. As previously described [9, 59], these data were fit using Sigma Plot to the Langmuir isotherm for cooperative two-state binding of two protein monomers to one DNA duplex:

$$\phi_{Fit} = 1/(1 + K_D^2/[P]^2). \quad (4)$$

[P] is the concentration of monomeric protein, and K_D is the dissociation constant. These fits yielded the apparent dissociation constant

K_D, from which the concentration of protein, $[P]_{1/2}$, where half-maximal DNA binding occurs, was calculated as $[P]_{1/2} = \sqrt{K_D}$.

Acknowledgments

We thank the University of Birmingham and the BBSRC for financial support and Baz Jackson for the use of the Luminescence Spectrometer.

Received: September 5, 2003
Revised: October 28, 2003
Accepted: October 28, 2003
Published: January 23, 2004

References

1. Allemann, R.K. (1999). DNA:protein binding specificity. In Encyclopedia of Molecular Biology, T.E. Creighton, ed. (New York: Wiley), pp. 743–744.
2. Murre, C., McCaw, P.S., and Baltimore, D. (1989). A new DNA binding and dimerisation motif in Immunoglobulin Enhancer binding, daughterless, MyoD, and *myc* proteins. Cell *56*, 777–783.
3. Molkentin, J.D., and Olson, E.N. (1996). Defining the regulatory networks for muscle development. Curr. Opin. Genet. Dev *6*, 445–453.
4. Sommer, L., Shah, N., Rao, M., and Anderson, D.J. (1995). The cellular function of MASH-1 in autonomic neurogenesis. Neuron *15*, 1245–1258.
5. Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D., and Weintraub, H. (1989). MyoD is a sequence-specific DNA binding protein requiring a region of *myc* homology to bind to the muscle creatine kinase enhancer. Cell *58*, 823–831.
6. Johnson, J.E., Birrin, S.J., and Anderson, D.J. (1990). Two rat homologues of drosophila *achaete-scute* specifically expressed in neuronal precursors. Nature *346*, 858–861.
7. Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronona, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell *66*, 305–315.
8. Kophengnavong, T., Michinowicz, J.E., and Blackwell, T.K. (2000). Establishment of distinct MyoD, E2A, and twist DNA binding specificities by different basic region-DNA conformation. Mol. Cell. Biol. *20*, 261–272.
9. Meierhans, D., el Ariss, C., Neuenschwander, M., Sieber, M., Stackhouse, J.F., and Allemann, R.K. (1995). DNA binding specificity of the basic helix-loop-helix protein MASH-1. Biochemistry *34*, 11026–11036.
10. Künne, A.G.E., Meierhans, D., and Allemann, R.K. (1996). Basic helix-loop-helix protein MyoD displays modest DNA binding specificity. FEBS Lett. *391*, 79–83.
11. Künne, A.G.E., and Allemann, R.K. (1997). Covalently linking bHLH subunits of MASH-1 increases specificity of DNA binding. Biochemistry *36*, 1085–1091.
12. Künne, A.G.E., Sieber, M., Meierhans, D., and Allemann, R.K. (1998). Thermodynamics of the DNA binding reaction of transcription factor MASH-1. Biochemistry *37*, 4217–4223.
13. Olson, E.N., Perry, M., and Schulz, R.A. (1995). Regulation of muscle differentiation by the MEF2 family of MADS box transcription factors. Dev. Biol. *172*, 2–14.
14. Mao, Z., and Nadal-Ginard, B. (1996). Functional and physical interactions between mammalian *Achaete-Scute* homolog-1 and myocyte enhancer factor 2A. J. Biol. Chem. *271*, 14371–14375.
15. Molkentin, J.D., Black, B.L., Martin, J.F., and Olson, E.N. (1995). Cooperative activation of muscle gene expression by MEF-2 and myogenic bHLH proteins. Cell *83*, 1125–1136.
16. Allemann, R.K. (1999). DNA:protein interaction thermodynamics. In Encyclopedia of Molecular Biology, T.E. Creighton, ed. (New York: Wiley), pp. 745–755.
17. Miroshnikov, A.I., Elyakova, E.G., Kudelin, A.B., and Senyavina, L.B. (1978). A study of the physicochemical characteristics of the neurotoxin apamin from the venom of the honeybee *Apis mellifica*. Bioorganicheskaya Khimiya *4*, 1022–1028.
18. Bystrov, V.F., Okhanov, V.V., Miroshnikov, A.I., and Ovchinnikov, Y.A. (1980). Solution spatial structure of apamin as derived from NMR study. FEBS Lett. *119*, 113–117.
19. Pease, J.H.B., and Wemmer, D.E. (1988). Solution structure of apamin determined by nuclear magnetic resonance and distance geometry. Biochemistry *27*, 8491–8499.
20. Pease, J.H.B., Storrs, R.W., and Wemmer, D.E. (1990). Folding and activity of hybrid sequence, disulfide-stabilized peptides. Proc. Natl. Acad. Sci. USA *87*, 5643–5647.
21. Brazil, B.T., Cleland, J.L., McDowell, R.S., Skelton, N.J., Paris, K., and Horowitz, P.M. (1997). Model peptide studies demonstrate that amphipathic secondary structures can be recognized by the chaperonin GroEL (cpn60). J. Biol. Chem. *272*, 5105–5111.
22. Ma, P.C.M., Rould, M.A., Weintraub, H., and Pabo, C.O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: Perspectives on DNA recognition and implications for transcriptional activation. Cell *77*, 451–459.
23. Cramer, J., Fiori, S., Müller, G., Renner, C., Pegoraro, S., and Moroder, L. (1999). Comarision of a Monte Carlo strategy with a combined DG/MDSA method for structure determination of bicyclic peptides. J. Mol. Model. *5*, 287–295.
24. Dempsey, C.E., Sessions, R.B., Lamble, N.V., and Campbell, S.J. (2000). The aparagine-stabilized β-turn of apamin: contribution to structural stability from dynamics simulation and amide hydrogen exchange analysis. Biochemistry *39*, 15944–15952.
25. Labbé-Jullié, C., Granier, C., Albericio, F., Defendini, M.L., Cearo, B., Rochat, H., and Van Rietcshoten, J. (1991). Binding and toxicity of apamin. Characterization of the active site. Eur. J. Biochem. *196*, 639–645.
26. Zimm, B.H., and Bragg, J.K. (1959). Theory of the phase transition between helix and random coil in polypeptide chains. J. Chem. Phys. *31*, 526–535.
27. Munoz, V., and Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. Nat. Struct. Biol. *1*, 399–409.
28. Greenfield, N., and Fasman, G.D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry *8*, 4108–4115.
29. Buskin, J.N., and Hauschka, S.D. (1989). Identification of a myocyte nuclear factor that binds to the muscle-specific enhancer of the mouse muscle creatine kinase gene. Mol. Cell. Biol. *9*, 2627–2640.
30. Weiel, J., and Hershey, J.W.B. (1981). Fluorescence polarisation studies of the interaction of *Escherichia coli* protein synthesis initiation factor 3 with 30S ribosomal subunits. Biochemistry *20*, 5859–5865.
31. Weiel, J., and Hershey, J.W.B. (1982). The binding of fluorescein-labeled protein synthesis initiation factor 2 to *Escherichia coli* 30S ribosomal subunits determined by fluorescence polarisation. J. Biol. Chem. *257*, 1215–1220.
32. Heyduk, T., and Lee, J.C. (1990). Application of fluorescence energy transfer and polarisation to monitor *Escherichia coli* cAMP receptor protein and *lac* promoter interaction. Proc. Natl. Acad. Sci. USA *87*, 1744–1748.
33. Ellenberger, T. (1994). Getting a grip on DNA recognition–structure of the basic region leucine zipper, and the basic region helix-loop-helix DNA binding domains. Curr. Opin. Struct. Biol. *4*, 12–21.
34. Brownlie, P., Ceska, T.A., Lamers, M., Romier, C., Stier, H., Teo, H., and Suck, D. (1997). The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. Structure *5*, 509–520.
35. Brennan, R.G., and Matthews, B.W. (1989). The helix-turn-helix DNA-binding motif. J. Biol. Chem. *264*, 1903–1906.
36. Harrison, S.C., and Aggarwal, A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. *59*, 933–969.
37. Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M., and Pabo, C.O. (1982). Homology among DNA binding proteins suggests use of a conserved super-secondary structure. Nature *298*, 447–451.

38. Pabo, C., and Lewis, M. (1982). The operator-binding domain of λ repressor: structure and DNA recognition. Nature *298*, 443–447.
39. Allemann, R.K. (1999). Transcription factors. In Encyclopedia of Molecular Biology, T.E. Creighton, ed. (New York: Wiley), pp. 2586–2593.
40. Montclare, J.K., and Shepartz, A. (2003). Miniature homeodomains: high specificity without the N-terminal arm. J. Am. Chem. Soc. *125*, 3416–3417.
41. Olson, E.N. (1990). MyoD family: a paradigm for development. Genes Dev. *4*, 1454–1461.
42. Paterson, B.M., Walldorf, U., Eldridge, J., Dübendorfer, A., Fratsch, M., and Gehring, W.J. (1991). The Drosophila homolog of vertebrate myogenic-determination genes encodes a transiently expressed nuclear protein marking primary myogenic cells. Proc. Natl. Acad. Sci. USA *88*, 3782–3786.
43. Zhang, R.G., Joachimiak, A., Lawson, C.L., Schevitz, R.W., Otwinowski, Z., and Sigler, P.B. (1987). The crystal structure of *trp* aporepressor at 1.8Å shows how binding tryptophan enhances DNA affinity. Nature *327*, 591–597.
44. Gryk, M.R., and Jardetzky, O. (1996). AV77 hinge mutation stabilises the helix-turn-helix domain of *trp* Repressor. J. Mol. Biol. *255*, 204–214.
45. Arvidson, D.N., Pfau, J., Hatt, J.K., Shapiro, M., Pecoraro, F.S., and Youderian, P. (1993). Tryptophan super-repressors with alanine 77 changes. J. Biol. Chem. *268*, 4362–4369.
46. Armstrong, A.P., Franklin, A.A., Uittenbogaard, M.N., Giebler, H.A., and Nyborg, J.K. (1993). Pleiotropic effect of the human T-cell leukemia virus Tax protein on the DNA binding activity of eukaryotic transcription factors. Proc. Natl. Acad. Sci. USA *90*, 7303–7307.
47. Wagner, S., and Green, M.R. (1993). HTLV-I Tax protein stimulation of DNA binding of bZip proteins by enhancing dimerisation. Science *262*, 395–399.
48. Baranger, A.M., Palmer, C.R., Hamm, M.K., Giebler, H.A., Brauweiler, A., Nyborg, J.K., and Schepartz, A. (1995). Mechanism of DNA binding enhancement by the human T-cell leukaemia virus transactivator Tax. Nature *376*, 606–608.
49. Perini, G., Wagner, S., and Green, M.R. (1995). Recognition of bZip proteins by the human T-cell leukaemia virus transactivator Tax. Nature *376*, 602–605.
50. Palmer, C.R., Gegnas, L.D., and Schepartz, A. (1997). Mechanism of DNA binding enhancement by Hepatitis B virus protein pX. Biochemistry *36*, 15349–15355.
51. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Jr., Fergusaon, D.M., Spellamyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. *117*, 5179–5197.
52. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD–visual molecular dynamics. J. Mol. Graph. *14*, 33–38.
53. Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. J. Mol. Biol. *189*, 113–130.
54. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.
55. Ellman, G.L. (1959). Tissue sulfhydryl groups. Arch. Biochem. Biophys. *82*, 70–77.
56. Mejbaum-Katzenellenbogen, S., and Drobryszycka, W.J. (1959). New method for quantitative determination of serum proteins separated by paper electrophoresis. Clin. Chem. Acta *4*, 515–522.
57. Lakowicz, J.R. (1999). Principles of Fluorescence Spectroscopy, Second Edition (New York: Plenum Press).
58. Heyduk, T., Ma, Y., Tang, H., and Ebright, R.H. (1996). Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. Methods Enzymol. *274*, 492–503.
59. Bird, G.H., Lajmi, A.R., and Shin, J.A. (2002). Sequence-specific recognition of DNA by hydrophobic, alanine-rich mutants of the basic region/leucine zipper motif investigated by fluorescence anisotropy. Biopolymers *65*, 10–20.