Molecular BioSystems

Dynamic Article Links

Cite this: *Mol. BioSyst.*, 2012, **8**, 1286–1296

www.rsc.org/molecularbiosystems

PAPER

Forced homodimerization of the c-Fos leucine zipper in designed bHLHZ-like hybrid proteins MaxbHLH-Fos and ArntbHLH-Fos†

Gang Chen,‡[a] Antonia T. De Jong[a] and Jumi A. Shin*[ab]

Received 29th August 2011, Accepted 10th January 2012
DOI: 10.1039/c2mb05354c

Although the c-Fos leucine zipper (LZ) does not form a homodimer in its native basic region/leucine zipper (bZIP) structure, we found that it is capable of homodimerization and promoting protein folding in engineered basic region/helix–loop–helix/leucine zipper (bHLHZ) hybrid proteins MaxbHLH-Fos and ArntbHLH-Fos, in which the bHLH subdomains of Max and Arnt are fused to the c-Fos LZ. By using the *in vivo* yeast one-hybrid system and *in vitro* circular dichroism and quantitative fluorescence anisotropy, we demonstrated that attachment of the c-Fos LZ to the otherwise unstructured MaxbHLH resulted in a hybrid bHLHZ-like protein now competent for homodimerization and DNA binding at the cognate E-box site, CACGTG. In ArntbHLH-Fos, the c-Fos LZ promoted proper folding of the HLH structure, although unlike MaxbHLH, ArntbHLH alone is capable of homodimerization and DNA binding. In addition, by comparing the E-box binding and secondary structures of MaxbHLH-Fos and two derivatives containing targeted mutations in the c-Fos LZ, we found that cooperative communication exists between the bHLH and LZ: proper folding of the four-helix bundle in the HLH region could be induced by the c-Fos LZ, and the HLH dimerization region could force homodimerization of the c-Fos LZ. These results demonstrate that although intrinsically unfavorable, the c-Fos LZ can homodimerize, demonstrating that the same c-Fos LZ element can yield orthogonal differences in structure and/or DNA-binding function within different transcription factor families, including the bZIP and bHLHZ.

Introduction

Transcription factors are a large and diverse class of DNA-binding proteins that target specific DNA sequences located in promoter or enhancer regions. They play a critical role in the transcriptional regulation of physiological functions including cell development, growth, and differentiation.[1] Thus, aberrant expression or incorrect processing of transcription factors (TFs) contributes to the progression of a variety of diseases, including developmental abnormalities and cancer.[2] The structurally simple α-helix scaffold is a common secondary structure utilized by TFs to bind the DNA major groove.[3] Structurally related motifs of DNA-binding proteins, including basic region/leucine zipper (bZIP), basic region/helix–loop–helix (bHLH), basic region/helix–loop–helix/leucine zipper (bHLHZ), and basic region/helix–loop–helix/Per-Arnt-Sim (bHLH/PAS), use α-helical scaffolds not only to target specific DNA sequences, but also to provide the protein dimerization interface, as all four motifs function as dimeric TFs.

The bZIP is the simplest structure of any known TF family, as it binds to the DNA major groove as a hetero- or homodimer of seamless α-helices.[4–7] Classic subdomain swapping experiments between bZIP proteins GCN4 and C/EBP confirmed that the bZIP is a modular structure comprising a basic region and a leucine zipper (LZ),[8,9] and that sequence-specific DNA-binding activity resides in the basic region, while dimerization specificity is determined by the leucine zipper.[10] Subdomain swapping between bZIP proteins c-Fos and GCN4 indicated that the LZ is interchangeable *within* the bZIP protein family and is responsible for dimerization partner specificity.[11]

The bHLH proteins also use a similar dimeric, helical DNA-binding basic region, but have a more complex HLH dimerization interface. The dimerization domain of bHLH proteins comprises two amphipathic helices separated by a nonconserved loop typically 5–12 residues in length; upon dimerization, a compact hydrophobic four-helix bundle forms that positions the contiguous basic regions for DNA binding.[12,13]

[a] *Department of Chemistry, University of Toronto, Mississauga, Ontario, Canada L5L 1C6. E-mail: jumi.shin@utoronto.ca; Fax: +1 905 828-5425; Tel: +1 905 828-5355*
[b] *Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada M5S 3G9*

† Electronic supplementary information (ESI) available: Circular dichroism spectroscopy of MaxbHLH-Fos and its two derivatives under different conditions, details of gene assembly and construction, details about Arnt isoforms. See DOI: 10.1039/c2mb05354c

‡ Current address: Terrence Donnelly Center for Cellular and Biomolecular Research and Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada M5S 3E1.

This journal is © The Royal Society of Chemistry 2012

While proteins with only the HLH as a dimerization interface exist, including MyoD and E47, other subfamilies utilize additional dimerization subdomains. The bHLHZ subfamily possesses a leucine zipper contiguous with the HLH subdomain.[14–16] The bHLH/PAS proteins use the globular Per-Arnt-Sim (PAS) homology domain as an additional dimerization interface, though PAS domains may have physiological functions other than dimerization.[17–20] Although not as self-evident as that within the bZIP family, subdomain swapping has also been shown to be feasible *between these TF families*.[18,21,22]

We designed bHLHZ-like hybrid proteins, MaxbHLH-Fos and ArntbHLH-Fos, that contain the c-Fos LZ and are capable of homodimerization and specifically targeting the E-box, in spite of the unfavorable Fos/Fos interaction.[23–25] The bZIP protein c-Fos heterodimerizes with c-Jun to bind the AP-1 site, 5′-TGACTCA,[26] and while c-Jun is capable of weak homodimerization, c-Fos is incapable of homodimerization.[23,24,27] The LZ of c-Fos is highly charged, making self-association electrostatically unfavorable. In comparison, Max and Arnt are capable of binding as homodimers to the Enhancer box (E-box, 5′-CACGTG). Max is a bHLHZ protein that can bind to the E-box as a homodimer or as the required heterodimerization partner of members of the Myc/Max/Mad TF network.[28] Similarly, Arnt is the required partner for other TFs in the bHLH/PAS subfamily. While Arnt is capable of binding to the E-box as a homodimer,[29–32] it can also bind to xenobiotic response elements (XRE, 5′-TNGCGTG) with binding partner AhR,[33,34] or to hypoxia response elements (HRE, 5′-RCGTG, where R is A or G) with binding partner HIF.[35]

In this work, we demonstrate the E-box-specific binding of hybrid proteins MaxbHLH-Fos and ArntbHLH-Fos by the *in vivo* yeast one-hybrid (Y1H) system and by *in vitro* circular dichroism (CD) and quantitative fluorescence anisotropy. Our results show that the c-Fos LZ is indeed involved in homodimerization in the MaxbHLH-Fos and ArntbHLH-Fos proteins, although homodimerization is intrinsically disfavored for the c-Fos LZ. We observe that in the context of the bHLHZ structural motif, the c-Fos LZ can be forced to form a homodimer.

Results

Since its discovery, the leucine zipper of c-Fos protein has been considered incapable of forming a homodimer.[23–25] However, in the development of artificial bHLHZ transcription factors, we were surprised to find that engineered protein ArntbHLH-Fos, in which the bHLH region of Arnt was fused to the LZ of c-Fos (alignment of hybrid proteins is shown in Fig. 1), was capable of E-box-responsive reporter gene activation in the Y1H assay; the extent of activation was not strong and somewhat variable, but readily detected and reproducible (Fig. 2 and Fig. S1, ESI†). Although c-Fos LZ homodimerization is unfavorable, we speculated that the c-Fos LZ must be involved in homodimerization of ArntbHLH-Fos, as repulsion between the two c-Fos LZ monomers would otherwise force dissociation of the Arnt HLH subdomains and lead to loss of DNA binding by the fusion protein.

We then tested whether the c-Fos LZ would behave similarly in the MaxbHLHZ domain, in which the native Max LZ is replaced by the c-Fos LZ, to probe whether this result was peculiar to Arnt; because Arnt is a member of the bHLH/PAS protein family, which lacks the leucine zipper element, we speculated that exchanging LZs in MaxbHLHZ might yield a different result. We also have found MaxbHLHZ to be more tractable than ArntbHLH; MaxbHLHZ and its derivatives form functional structures more reliably than Arnt in yeast.[21,22] Both Brennan's and Whitelaw's groups reported that the dimerization and DNA-binding of the bHLH domain of Arnt are particularly sensitive to ionic strength, most likely because it has a more hydrophilic dimerization interface than those in other bHLH proteins.[36,37] Thus, the E-box-responsive activity displayed by ArntbHLH-Fos may be specific for Arnt's unusual structural properties.

We therefore constructed hybrid protein MaxbHLH-Fos, with the c-Fos LZ replacing the native LZ, and similarly tested its E-box-responsive activation in the Y1H. In this case, we found that MaxbHLH-Fos strongly activated the E-box-responsive reporter gene in the Y1H. This prompted us to further explore the potential homodimerization of the c-Fos LZ in non-native bHLHZ hybrid proteins.

```
                     basic region          helix-loop-helix                    leucine zipper
                  82
ArntbHLH-Fos      SSADKERLARENHSEIERRRR-NKMTAYITELSDMVPTCSALARKPDKLTILRMAVSHMKSL-RGTGNTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRP
ArntbHLH       DQMSNDKERFARENHSEIERRRR-NKMTAYITELSDMVPTCSALARKPDKLTILRMAVSHMKSL-RGT
ArntbHLH-C/EBP DQMSNDKERFARENHSEIERRRR-NKMTAYITELSDMVPTCSALARKPDKLTILRMAVSHMKSL-RGTRIRLEQKVLELTSDNDRLRKRVEQLSRELDTL
                                                                                  166                                202
Fos LZ                                                                            QAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRP
                     22
MaxbHLHZ             ADKRAHHNALERKRR-DHIKDSFHSLRDSVPSLQGEKAS  RAQILDKATEYIQYM-       RRKNDTHQQDIDDLKRQNALLEQQVRALEKARSSAQLQT
MaxbHLH-Fos          ADKRAHHNALERKRR-DHIKDSFHSLRDSVPSLQGEKAS  RAQILDKATEYIQYM-       QAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRP
MaxbHLH-FosLA        ADKRAHHNALERKRR-DHIKDSFHSLRDSVPSLQGEKAS  RAQILDKATEYIQYM-       QAETDQAEDEKSAAQTEIANALKEKEKAEFILAAHRP
MaxbHLH-FosW         ADKRAHHNALERKRR-DHIKDSFHSLRDSVPSLQGEKAS  RAQILDKATEYIQYM-       QAEIEQLEERNYALRKEIEDLQKQLEKLGAP
```

Fig. 1 Sequence alignment of Arnt and Max bHLHZ-like hybrid proteins based on ref. 14, 15 and 36. The bHLH region of Max was aligned with the putative bHLH region of Arnt based on sequence similarity. Hyphens separate the basic region, HLH, and leucine zipper regions; spaces are used to maintain sequence alignments. ArntbHLH-Fos and the previously reported ArntbHLH and ArntbHLH-C/EBP[21] differ in that the bHLH of the ArntbHLH and ArntbHLH-C/EBP constructs has two extra residues (Asp, Gln) at the N terminus, plus three different residues (M82S, N84A, and F89L). Compared with ArntbHLH and ArntbHLH-C/EBP, ArntbHLH-Fos was designed based on a different Arnt isoform (see the Experimental section for details). As these residues are located upstream of the putative DNA-binding basic region, these differences were not expected to affect the comparison between ArntbHLH-Fos and the other two Arnt constructs. The three conserved residues in the Max basic region that make sequence-specific contacts to the DNA major groove, as well as the corresponding residues in the Arnt basic region, are in bold, italic.[14,15] The leucines/hydrophobic residues that define the leucine zipper motif are in bold. The mutated residues in MaxbHLH-FosLA and MaxbHLH-FosW are shown in italic, underlined. The three RIR residues that are italicized in the LZ region of ArntbHLH-C/EBP encode the BamH I restriction site for the purpose of cloning.

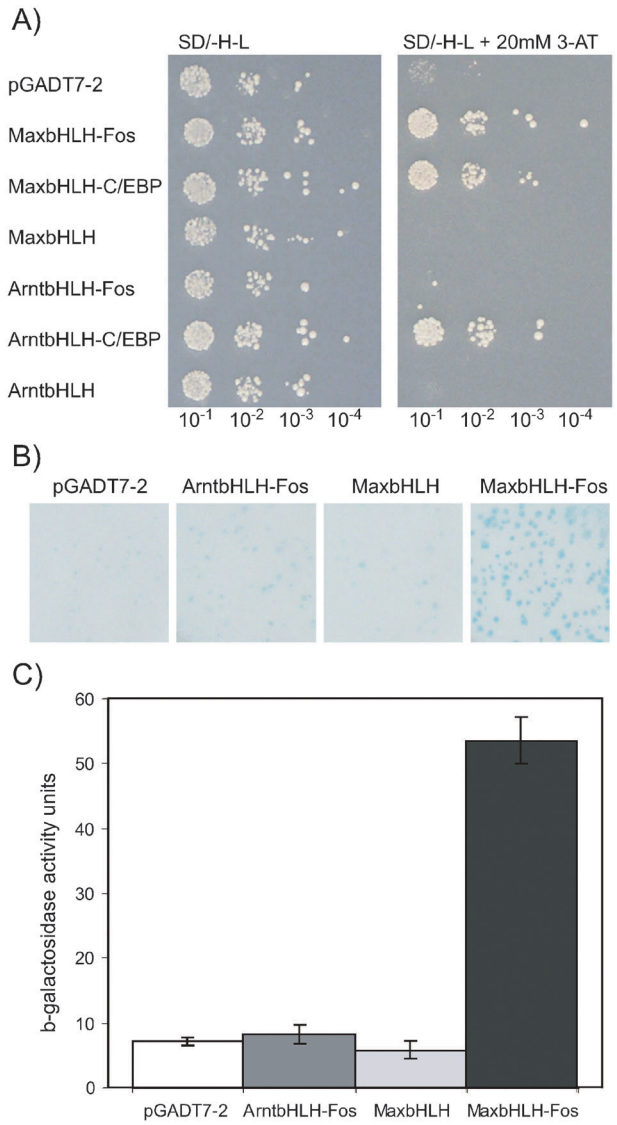

Fig. 2 E-box-responsive transcriptional activation of ArntbHLH-Fos and MaxbHLH-Fos in Y1H. (A) The *HIS3* assay. Vectors pGADT7-2 expressing different proteins as C-terminal fusions to GAL4 AD were transformed into yeast strain YM4271[pHISi-1/E-box]. The resulting yeast colonies expressing these GAL4 AD fusions are listed along the left, with control expressing GAL4 AD only at the top. Colonies were spotted (10 μL) as ten-fold serial dilutions onto SD/-His/-Leu medium to select for the presence of plasmid, followed by spotting on SD/-His/-Leu medium with 20 mM 3-AT to select for DNA-binding protein leading to colony growth. Data for a second independent trial for MaxbHLH-Fos, ArntbHLH-Fos, and MaxbHLH are provided in Fig. S1, ESI.† (B) The X-gal colony-lift filter assay. Yeast strain YM4271[pLacZi/E-box] was transformed with vectors pGADT7-2 expressing four different GAL4 AD fusion proteins as listed. The resulting transformants were incubated on SD/-Ura/-Leu plates at 30 °C, 4 days before testing. Photos were taken after 2 hours incubation. (C) Histogram comparing binding strengths of each hybrid protein to E-box. All values are averages of 9–12 measurements (±SEM) from 3–4 independent measurements conducted in triplicate.

Fos LZ-containing bHLHZ hybrids produce E-box-responsive gene activation in Y1H assay

In the Y1H assay, two reporter genes, *HIS3* and *lacZ*, were independently used to test the E-box-responsive activity of the two hybrid proteins. In the spot titration assay, where serial dilutions of each transformant were spotted on both control and test plates, activation of the *HIS3* reporter was observed for both ArntbHLH-Fos and MaxbHLH-Fos on the test plate with 20 mM 3-aminotriazole inhibitor (3-AT, Fig. 2). MaxbHLH, with no LZ appended to the HLH, displayed no reporter activation. We compare these results with previously generated ArntbHLH-C/EBP and MaxbHLH-C/EBP, in which the bHLH domains of Arnt and Max were fused to the LZ from bZIP transcription factor C/EBP; in contrast to the c-Fos LZ, the C/EBP LZ is a strong homodimerization element.[38] Both ArntbHLH-C/EBP and MaxbHLH-C/EBP showed reporter gene activation in the *HIS3* reporter assay, whereas the ArntbHLH alone produced negative signals.[21] Although negative results from the Y1H assay do not necessarily demonstrate that the protein tested is incapable of reporter gene activation under all circumstances, the positive signals from ArntbHLH-Fos and MaxbHLH-Fos indicate that specific binding at the E-box site was most likely involved, thereby triggering reporter gene activation.

The two *lacZ*-based β-galactosidase assays, the quantitative ONPG assay and the qualitative but more sensitive X-gal colony-lift filter assay, were further utilized to confirm the results obtained from the *HIS3* assay. In the colony-lift assay, ArntbHLH-Fos and MaxbHLH-Fos showed different transcriptional activation potencies from the E-box-responsive *lacZ* reporter gene (Fig. 2). Blue color appeared for MaxbHLH-Fos within 30 minutes, becoming bright blue after 1 hour. In comparison, the control MaxbHLH did not develop detectable blue color after 1 hour. Blue color only slightly more intense than the background appeared after 2 hours for ArntbHLH-Fos, similar to MaxbHLH. The ONPG data confirmed the results from the colony-lift assay. The ONPG values for ArnbHLH-Fos and MaxbHLH were statistically the same as for the negative control pGADT7-2 (8 ± 1 and 6 ± 1 *vs.* 7 ± 1, respectively). In comparison, MaxbHLH-Fos gave a positive ONPG reading of 54 ± 4, clearly much higher than the background.

Quantitativve fluorescence anisotropy measurements of protein:DNA complexation confirm Y1H results

In vitro quantitative fluorescence anisotropy titrations were performed to measure dissociation constants for protein:E-box complexation. No binding by any protein was detected with the nonspecific DNA control, even at 2 μM, the highest monomeric protein concentration tested (data not shown). Both ArntbHLH-Fos and MaxbHLH-Fos specifically targeted the E-box site (Fig. 3). Analyses of E-box-binding isotherms showed that both ArntbHLH-Fos and MaxbHLH-Fos, as well as designed mutants MaxbHLH-FosLA and MaxbHLH-FosW that are discussed below, gave Hill coefficient values similar to that of native MaxbHLHZ.[22] Therefore, these proteins, like MaxbHLHZ, likely form dimeric structures for cooperative binding to the E-box site.

The dissociation contant (K_d) of the ArntbHLH-Fos complex with the E-box is 436 ± 107 nM, which is much weaker than the previously reported K_d values for ArntbHLH and ArntbHLH-C/EBP binding to E-box (40.2 ± 10.7 nM and 148.9 ± 2.9 nM, Table 1).[21] This result confirms two expected

This journal is © The Royal Society of Chemistry 2012

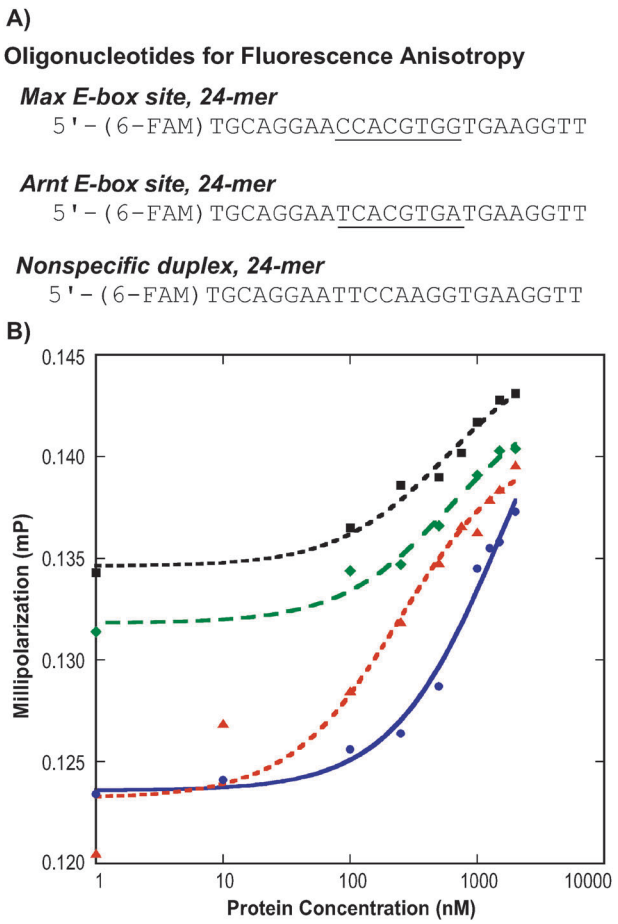

Fig. 3 Fluorescence anisotropy. (A) DNA duplexes used in fluorescence anisotropy titrations. "6-FAM" is 6-carboxyfluorescein. Max-preferred and Arnt-preferred E-boxes are underlined and share the core E-box, CACGTG. (B) Representative equilibrium binding isotherms for ArntbHLH-Fos (■, black dashed line), MaxbHLH-Fos (◆, green dashed line), MaxbHLH-FosLA (●, blue solid line) and MaxbHLH-FosW (▲, red dashed line) targeting the E-box. Each isotherm was obtained from an individual titration, and each K_d value is the average of two individual titrations ± SEM. Titrations were performed in either Tris buffer (ArntbHLH-Fos) or phosphate buffer (Max derivatives, see the Experimental section for details).

conclusions: first, the Arnt bHLH domain itself is able to target the E-box DNA site without the aid of the secondary dimerization domain PAS, as reported previously.[36,37] Second, although the LZ is not necessary for E-box binding, an LZ that favors homodimerization (*i.e.* C/EBP LZ in ArntbHLH-C/EBP) produces a stronger DNA-binding TF than a LZ that disfavors homodimerization (*i.e.* c-Fos LZ in ArntbHLH-Fos).

These observations for the ArntbHLH fusions to either the c-Fos or C/EBP LZ are also supported by our findings for the MaxbHLH fusions to the same LZs. Although no dissociation constants were measured for the previously reported MaxbHLH-C/EBP hybrids, comparison of the ONPG values of MaxbHLH-Fos (54 ± 4) and the MaxbHLH-C/EBP hybrids (95.3 ± 4.0, 102.6 ± 9.5, and 86.2 ± 5.0, for the Met, Thr, and Ile hybrids, respectively)[39] demonstrates that E-box-responsive activation is higher with the C/EBP LZ, but that the c-Fos LZ still contributes positively to activation, because the MaxbHLH itself shows no activity in Y1H assays and fluorescence anisotropy titrations. However, none of these Max hybrids approach the ONPG value measured for the native MaxbHLHZ (147.4 ± 7.3).[22] Finally, the lack of E-box-binding by MaxbHLH confirms that unlike ArntbHLH, the bHLH domain of Max requires assistance from the secondary LZ dimerization subdomain for specific DNA binding.

MaxbHLH-Fos mutants further demonstrate the c-Fos LZ role in dimerization and E-box binding

To further assess the contributions of the c-Fos LZ element toward protein homodimerization and DNA binding, two mutants based on hybrid MaxbHLH-Fos were designed. In MaxbHLH-FosLA, all leucine residues in d positions of the α-helix heptad were mutated to alanine in the c-Fos LZ. These Leu-to-Ala mutations in the LZ homodimer should significantly decrease, if not abolish, the stability of the hydrophobic core that stabilizes the homodimeric coiled-coil structure, whereas the electrostatic repulsive forces still exist. However, the α-helical structure of the monomeric LZ should be maintained, if not improved, as Ala is the strongest α-helix former and stabilizer of the natural amino acids.[40] Therefore, if the LZ is simply serving as a structural device that nucleates and stabilizes the α-helical protein fold, the MaxbHLH-FosLA mutant can test this hypothesis. The loss of the LZ hydrophobic core in MaxbHLH-FosLA was anticipated to disrupt homodimerization and DNA binding. As expected, the dissociation constant of the MaxbHLH-FosLA complex with the E-box is over 1 μM, an affinity ≥ two-fold weaker than that of MaxbHLH-Fos (K_d value 536 ± 83 nM, Table 1 and Fig. 3).

In contrast, MaxbHLH-FosW, in which the c-Fos LZ was replaced with FosW,[27] showed significantly increased ability to bind the E-box, with a K_d value of 246 ± 4 nM (Table 1 and Fig. 3). FosW is a mutant of the c-Fos LZ identified as not only a heterodimerization partner for c-Jun, but also FosW is capable of homodimerization. In FosW, the two Lys residues in the c-Fos LZ core that prevent homodimerization were mutated to Asn and Ile, respectively, which stabilize FosW homodimerization (Fig. 1).[41,42] Replacement of Lys with Asn in position a of the heptad is believed to be critical in maintaining partner specificity.[41] In addition, a Thr-to-Ile mutation in position a of the heptad favors homodimerization.[43]

Taken together, these *in vitro* data indicate that the leucines in d positions of the c-Fos LZ contribute positively toward stabilizing the dimerization interface of the c-Fos homodimer, and that when MaxbHLH-Fos is bound to the E-box site, the hydrophobic interfaces presented by the Max HLH and the Leu-rich LZ are sufficient to overcome the intrinsic charge repulsion between the two c-Fos LZs. This is in contrast to the c-Fos LZs in the *native bZIP structure*: as the sole dimerization element in the bZIP, the charge repulsion would outcompete the attractive hydrophobic interaction and lead to dissociation of the two c-Fos monomers.

Moreover, the decrease (in the case of MaxbHLH-FosLA) or increase (in the case of MaxbHLH-FosW) in DNA-binding affinity also indicates a significant difference between the structures of the bHLHZ and bZIP regarding dimerization, which affects their DNA-binding functions. In bHLHZ proteins

Table 1 Helicities and DNA-binding affinities of bHLHZ hybrids

Protein	Helicity[a]				K_d (nM)[b], bound to E-box DNA
	0.5 μM protein, no DNA	2 μM protein, no DNA	2 μM protein + nonspecific DNA	2 μM protein + Max E-box DNA	
ArntbHLH[c]	—	~0%	—	—	40 ± 11[e]
ArntbHLH-C/EBP[c]	—	56%	—	—	149 ± 3[e]
ArntbHLH-Fos	—	21%	—	—	436 ± 107[e]
MaxbHLHZ[d]	16%	39%	40%	44%	14 ± 8
MaxbHLH	—[g]	~0%	—[g]	—[g]	> 2000[f]
MaxbHLH-Fos	15%	31%	58%	59%	536 ± 83
MaxbHLH-FosLA	~0%	16%	18%	45%	1400 ± 200
MaxbHLH-FosW	40%	55%	67%	74%	246 ± 4

[a] Helicity was measured by CD at 222 nm in the absence or presence of nonspecific or Max E-box DNA duplex. [b] Dissociation constants were measured by fluorescence anisotropy. See the Experimental section for details. [c] Value reported previously by Chow *et al.*[21] [d] Value reported previously by Xu *et al.*[22] [e] Measured in Tris buffer. See the Experimental section for details. [f] No binding was detected between 2 μM MaxbHLH and Max E-box DNA. [g] Because no binding to DNA was detected by MaxbHLH in either Y1H or fluorescence anisotropy titrations, CD was not performed at a lower protein concentration of 0.5 μM or in the presence of DNA. —, not determined.

MaxbHLH-Fos and the LA and W mutants, the HLH plays the essential role in protein dimerization:[44] the LZ—c-Fos, FosW, or FosLA—plays a secondary role in dimerization. The primary function of the LZ in bHLHZ transcription factors is partner selection.[16] In this work however, all our Max and Arnt hybrids were examined for homodimerization only; therefore, the LZ element in our hybrids can fine-tune the affinity of the protein/protein homodimeric interaction, and potentially dictate partner specificity in a pool of heterodimerization candidates. In the bZIP structure, the LZ is the only element responsible for dimerization and partner selection, and therefore, no other structural element plays a supporting role in the protein/protein interaction.

Circular dichroism demonstrates that c-Fos LZ element promotes folding of both ArntbHLH and MaxbHLH

We used circular dichroism (CD) to assess secondary structure of our engineered proteins. Because of the buffer compatibility issue discussed in the Experimental section, we used different buffer systems to measure DNA binding (fluorescence anisotropy) or protein structure (CD) of Arnt derivatives. With a short LZ, whether C/EBP or c-Fos, attached to the bHLH subdomain of Arnt, the hybrid proteins display helical structure much higher than that displayed by the ArntbHLH domain itself (no clear helical secondary structure[21]), indicating that the LZ promotes proper folding of protein structure (Table 1). Fused to the ArntbHLH, the C/EBP LZ appears to be more effective at nucleating helicity than the c-Fos LZ: 56% *vs.* 21% at 2 μM protein, respectively.

The native ArntbHLH is actually a strong binder of the E-box under our chosen *in vitro* conditions (K_d value 40 nM, see the Experimental section for conditions),[21] indicating that ArntbHLH is likely more folded under these conditions. Fusion of an additional LZ to the ArntbHLH makes it a weaker binder of the E-box: even ArntbHLH-C/EBP, with the additional strongly dimerizing C/EBP LZ, is a three-fold weaker binder of the E-box, and ArntbHLH-Fos is a ten-fold weaker binder, likely due to the repulsive nature of the c-Fos LZ (Table 1). So although the C/EBP and c-Fos LZs contribute to helical, folded structure, they do not contribute positively to binding function at the E-box site; rather, the dimerization potency of the LZ fine-tunes the binding function of Arnt derivatives at the E-box site in *in vitro* fluorescence anisotropy titrations. Interestingly, this *in vitro* result is opposite to that observed *in vivo* in the Y1H assay: no E-box-responsive reporter activation was observed for ArntbHLH alone.[21] We surmised that the yeast environment does not promote proper structure in ArntbHLH that will lead to E-box binding function.

In contrast to the ArntbHLH hybrids above, the Max derivatives show good correlation between helical structure and DNA-binding function. In the absence of DNA, MaxbHLH shows no helical structure (~0%, Table 1), and no detectable E-box-binding function was observed for MaxbHLH in any DNA binding assay. Appending the LZ to the MaxbHLH, however, now generates proteins with more helical structure and measurable E-box-binding function, indicating that a secondary dimerization element or folding-enhancing element is required to assist MaxbHLH in forming proper structure leading to DNA-binding function. Given that Max is a member of the bHLHZ protein family, the necessity of a LZ toward helical structure and DNA-binding function could be anticipated; in comparison, Arnt is a member of the bHLH/PAS family, so substitution of the PAS domain with a non-native LZ may not afford the same benefits toward structure and function as for Max.

The two MaxbHLH-Fos mutants displayed different levels of folded structure. The differences in MaxbHLH-Fos and MaxbHLH-FosW helicities (31% and 55% at 2 μM protein, Table 1) correlate well with the helicities of the c-Fos and FosW leucine zippers (17.3% and 43.7%), albeit measured by Worrall and Mason at 150 μM protein,[45] a concentration that promotes homodimerization, and much higher than our measurements at 0.5 and 2 μM (Table 1). The authors' short peptides were designed with N- and C-capping motifs to improve solubility, and unlike our constructs, relatively high concentrations were achievable without encountering problems with solubility. In spite of this extremely high protein concentration, the clear trend follows that the FosW LZ is markedly more helical than the c-Fos LZ under the same conditions.[45]

Interestingly, when all leucines in d positions of the heptads in the LZ of MaxbHLH-Fos are replaced with alanines, the resultant MaxbHLH-FosLA protein showed much less secondary structure

 This journal is © The Royal Society of Chemistry 2012

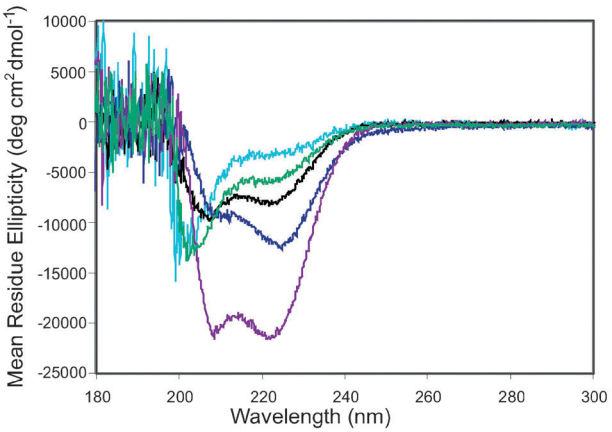

Fig. 4 Circular dichroism spectra (from top to bottom) of MaxbHLH (light blue), MaxbHLH-FosLA (green), ArntbHLH-Fos (black), MaxbHLH-Fos (dark blue), and MaxbHLH-FosW (purple) at 2 μM protein concentration. CD spectral overlays of the proteins in the absence of DNA or in the presence of nonspecific or Max E-box DNA are shown in Fig. S2 (ESI†), and spectra for proteins at 0.5 μM are shown in Fig. S3 (ESI†). Each spectrum was averaged twice, and curves were not subjected to smoothing. The buffer control was subtracted from each protein spectrum. Mean residue ellipticities are presented, which account for differences in lengths of proteins.

relative to MaxbHLH-Fos and MaxbHLH-FosW (16% *vs.* 31% and 55% at 2 μm protein, Table 1 and Fig. 4). This is likely due to removal of the hydrophobic core originally formed by the Leu residues in heptad d positions; without the stabilizing effects from the hydrophobic core, charge repulsion becomes the dominant force that destabilizes the homodimeric structure. In other words, the four-helix bundle of MaxbHLH cannot be stably formed without the support of a secondary dimerization domain.

These results also show correlation between the amount of helical secondary structure of the free proteins and the K_d values in the presence of E-box DNA in the case of Max derivatives: the trend from lowest to highest helicity in MaxbHLH-FosLA (16%), MaxbHLH-Fos (31%), and MaxbHLH-FosW (55%) is mirrored in their respective binding affinities for the E-box (1400 nM, 536 nM, and 246 nM). This trend may be explained by free-energy changes, especially the entropic contribution, to go from disordered to more ordered structure upon protein:DNA complexation.

Helicities of Max proteins were measured at both 0.5 μM (Fig. S3 (ESI†), Table 1) and 2 μM (Fig. 4 and Table 1) monomeric concentrations. All proteins showed an increase in helicity from 0.5 μM to 2 μM, indicating that the transition from the monomer to the dimer could occur to give the more helical coiled-coil structure. This correlates well with the dimerization dissociation constant of 622 nM measured for MaxbHLHZ at 23 °C,[46] and supports the likelihood that the monomer–dimer equilibrium constants of our Max proteins are similar in the high nanomolar–low micromolar range. At lower concentrations, more of the protein will be present in the monomeric form, which is typically less structured for bHLH proteins.[47–49]

The correlation of increase in helicity with a shift in monomer–dimer equilibrium is also supported by the work of Worrall and Mason.[45] Using circular dichroism, the authors analyzed helicity ratios $\Theta_{222}/\Theta_{208}$, which can describe the likelihood that a protein helix is formed in isolation (monomer) or within a coiled-coil structure (dimer). Their calculations suggest that the difference in helicities between the isolated c-Fos and FosW LZ elements correlates with the dimerization state of the zippers, with the less helical c-Fos LZ being present primarily as monomers, whereas the more structured FosW zipper suggests a coiled-coil structure.[45]

This observation indicates that inter-helical interaction must exist between the two c-Fos LZs, which leads to proper folding of the hybrid protein and DNA binding. Upon addition of DNA, protein helicity increases (Table 1 and Fig. S2, ESI†). In the presence of nonspecific DNA, both MaxbHLH-Fos and MaxbHLH-FosW become more helical (31% and 58% without DNA, *vs.* 55% and 67% with nonspecific DNA, respectively, Table 1). In the presence of the E-box DNA target, MaxbHLH-FosW shows a modest increase in helicity over that in the presence of nonspecific DNA (74% with E-box *vs.* 67% with nonspecific); MaxbHLH-Fos shows the same level of helical structure with nonspecific or E-box DNA (58% or 59%). In comparison, the structure of the weakly ordered MaxbHLH-FosLA (16%) did not change in the presence of nonspecific DNA (18%). This is in sharp contrast to the dramatic increase in helicity for the same protein in the presence of the E-box DNA target (45%), likely due to induced folding from DNA binding. This suggests that the lack of support from a secondary dimerization element can be compensated by another dimerizing factor—the bound DNA target. In the case of MaxbHLH-FosLA, it is likely that the interaction with nonspecific DNA is not energetically favorable enough to compensate for the entropic cost of ordering the structure of the weakly helical MaxbHLH-FosLA protein. Interaction with the specific Max E-box DNA should be more favorable, compensating for the loss of entropy.

Taken together, our structural CD results and quantitative DNA-binding function analyses indicate that the c-Fos LZ is capable of homodimerization and promoting formation of helical, folded structure in non-native bHLHZ hybrid proteins. As a secondary dimerization element appended to the primary dimerization element HLH, the c-Fos LZ contributes to protein folding, and in some cases homodimerization, in a supporting fashion.

Discussion

LZs can act as trigger sequences for initiation of four-helix bundle formation in bHLH

Although the existence of protein trigger sequences that promote coiled-coil formation remains controversial,[50] bHLHZ proteins can use the secondary dimerization element—a helical LZ—to promote dimerization of the HLH domain to form a four-helix bundle, regardless of the effect of the LZ on DNA-binding function. For example, in the case of MaxbHLH-Fos, the stable four-helix bundle of MaxbHLH could not be formed without the LZ; with ArntbHLH, CD also indicated that the presence of the LZ promoted folding of ArntbHLH (~0% helicity for ArntbHLH *vs.* 56% and 21% for ArntbHLH-C/EBP and ArntbHLH-Fos at 2 μM protein, respectively).

Trigger sequences do not always contribute to DNA binding. Chapman-Smith *et al.* showed that the ArntbHLH alone,

without the aid of the PAS region, is capable of forming homodimers that bind to the E-box site, although the PAS may increase protein stability.[37] We found that attachment of a non-native LZ to ArntbHLH considerably decreased its E-box binding *in vitro*, as in the case of ArntbHLH-Fos in this study and the previously reported ArntbHLH-C/EBP,[21] although the LZ does promote helical secondary structure, as shown by CD and supported by the necessity of the LZ for E-box-responsive reporter activation *in vivo*. The helical LZ in MaxbHLHZ can also serve as a trigger scaffold around which the remainder of the coiled-coil structure can "zip up."[50] For both ArntbHLH and MaxbHLH, fusion of a non-native LZ with higher dimerization potential generally leads to more well-folded structures.

The c-Fos LZ-containing protein is capable of homodimerization

The homodimer of the c-Fos leucine zipper itself is unstable and does not form in its native bZIP structure or as the isolated LZ.[23,24,27] The c-Fos LZ, however, readily heterodimerizes with the c-Jun LZ. This heterodimer is further stabilized by DNA binding, and its favorable energetics arise from intermolecular hydrophobic and electrostatic interactions.[26] Unlike the interaction between the LZs of bZIP proteins c-Jun and c-Fos, which are structurally more rigid with the protein/protein interface well matched to form the heterodimeric Jun/Fos coiled coil, homodimerization of the c-Fos LZ in the more flexible bHLHZ-like hybrid proteins MaxbHLH-Fos and ArntbHLH-Fos is a *forced* dimerization, due to the homodimerizing capability of the HLH domain adjacent to the LZ.

The interactions involved in the MaxbHLH-Fos and ArntbHLH-Fos homodimers are more complicated. If the interactions involved are dissected, they can be simplified by considering only the attractive hydrophobic interactions between the hydrophobic interfaces of the two LZs and the repulsive electrostatic interactions between the like-charged e and g′ residues in the neighboring helices forming the coiled coil. The c-Fos bZIP is incapable of homodimerization, indicating that the repulsion between like-charged residues is sufficient to overcome any attractive interactions at the hydrophobic interface and potentially attractive electrostatic interactions, as well as possible dimer-promoting contributions from two bZIP proteins binding the DNA target. In contrast, once the c-Fos LZ is fused to the bHLH domain of Max or Arnt, the attractive hydrophobic interactions and/or electrostatic interactions of the HLH subdomains, as well as contributions from E-box binding, are strong enough to promote formation of the four-helix bundle and force the two repulsive c-Fos LZs into close proximity.

The LZ region is required structurally and functionally for MaxbHLH, but not for ArntbHLH

The fact that the entire bHLHZ region is necessary for efficient dimerization and DNA binding by bHLHZ transcription factors has already been reported.[51,52] Our observations indicate that the difference in DNA-binding activities between MaxbHLH and ArntbHLH lies in the structural differences of their four-helix bundles, which should be similar, although no high-resolution structure exists for the ArntbHLH region. Max and Arnt belong to different subfamilies—bHLHZ and bHLH/PAS, respectively. The bHLH/PAS motif may possess subtle structural differences in the bHLH region that allow the ArntbHLH to bind strongly to the E-box target, without a secondary PAS or LZ element. In addition, comparison of the sequences of bHLH regions from bHLH-containing proteins shows that the loop regions vary significantly in composition and length.[12,14] These variations may impose structural differences in the loops, and therefore in the bHLH regions overall, that lead to different homodimerization capabilities.

The roles that the c-Fos zipper play in MaxbHLH-Fos and ArntbHLH-Fos are two-fold: the structural role is similar, as the fusion of the c-Fos LZ to the bHLH domain of Max or Arnt nucleates the α-helical structure that encourages proper folding of the HLH and solubility, which enhances protein stability and promotes homodimerization. The functional role of the c-Fos LZ fusion, however, differs. In Arnt, the bHLH domain itself is capable of homodimerization and high affinity binding to the E-box.[36] However, ArntbHLH is relatively intractable and prone to loss of function due to misfolding and/or aggregation under diverse conditions.[18,21,37] Therefore, the c-Fos LZ in ArntbHLH-Fos mainly acts as a folding- or solubility-enhancing tag. Regarding DNA-binding function, the K_d value of E-box binding by ArntbHLH-Fos is approximately ten-fold higher than its parental counterpart ArntbHLH and three-fold higher than that of its hybrid counterpart ArntbHLH-C/EBP,[21] in which the homodimer-promoting C/EBP LZ is fused to ArntbHLH (Table 1). In the case of ArntbHLH-Fos, our results point to the intrinsic charge repulsion of the c-Fos LZ being the most dominant effect, outweighing any positive contributions from the LZ hydrophobic core or increased protein folding and helicity, toward the decrease observed in DNA-binding function.

Unlike the bHLH domain of Arnt, the bHLH domain of Max is incapable of homodimerization without assistance from a contiguous leucine zipper at its C-terminus. Neither the MaxbHLH nor c-Fos LZ is competent for homodimerization or DNA binding. By stitching these two elements together, however, the resultant hybrid bHLHZ protein MaxbHLH-Fos is able to homodimerize and specifically target the E-box DNA duplex, as demonstrated by both *in vivo* Y1H assay and *in vitro* fluorescent anisotropy measurements.

Our results indicate that in MaxbHLH-Fos, the c-Fos LZ serves to enhance the proper folding of MaxbHLH to achieve the helical HLH structure necessary to achieve the four-helix bundle in the dimer and compensate for intrinsic repulsion between c-Fos zippers. Therefore, this communication between MaxbHLH and c-Fos LZ makes their fused structure competent for homodimerization and DNA-binding function.

The sequence differences between ArntbHLH-Fos and the previously reported ArntbHLH and ArntbHLH-C/EBP[21] occur outside of the DNA-binding basic region (see Fig. 1 and the caption for sequence alignment and description of amino-acid differences). Specifically, the three residues at the ArntbHLH N-terminus differ, and two residues before the basic region differ; a different Arnt isoform was used to construct ArntbHLH-Fos in comparison with the previous derivatives (see the Experimental section for details). This change does not appear to affect the comparison between ArntbHLH-Fos and the two previous Arnt constructs, and was not anticipated to do so, as these changes are

 This journal is © The Royal Society of Chemistry 2012

beyond the DNA-binding region. Rather, the results from the Arnt proteins are echoed by the results from the Max hybrids, discussed above.

DNA binding function does not strictly correlate with folded protein structure

Although MaxbHLH-Fos and its two mutants follow the trend that more folded helical structure correlates with higher E-box binding affinity, DNA-binding function does not strictly correlate with helical protein structure. For instance, the native MaxbHLHZ shows the highest E-box-binding affinity by far, yet MaxbHLHZ neither shows significantly ordered structure in its high-affinity complex with its cognate E-box target (only 44% helicity, Table 1) nor does the helicity of MaxbHLHZ increase significantly when bound to E-box (39% helicity without the presence of DNA). Additionally, the helical content of MaxbHLHZ is only modestly above that of MaxbHLH-Fos (39% *vs*. 31%) and markedly less than that of MaxbHLH-FosW (55%). This suggests that the c-Fos LZ is almost as efficient as the native Max LZ in the role of promoting correctly folded HLH structure. Thus, one might consider that the lower DNA-binding affinity of the hybrid proteins stems from the repulsive nature of the c-Fos LZ element. However, even using a LZ with stronger dimerizing potency does not yield a hybrid with DNA-binding capability comparable to the native protein: we reported that a set of MaxbHLH-C/EBP hybrid proteins shows significantly weaker ability to activate transcription from E-box-responsive reporters than does native MaxbHLHZ in the Y1H,[39] although the C/EBP LZ is a stronger homodimerization element than the Max LZ.[53]

We hypothesize that co-evolution of the Max LZ and bHLH domains may have lead to cooperative interactions between both subdomains resulting in an optimized, more flexible bHLHZ structure, whereas the C/EBP and c-Fos LZs were specifically co-evolved in the context of the more rigid bZIP structure. Therefore, although the non-native Fos LZ may serve comparably to the native Max LZ as a nucleation device to assist the MaxbHLH domain in forming secondary structure, the DNA binding of the non-native hybrid proteins is less efficient than that of the native MaxbHLHZ.

Additionally, we note that other factors, including non-specific or genomic DNA and buffer conditions, can influence protein folding and DNA-binding function; indeed, the folding and DNA binding of ArntbHLH are buffer dependent.[21,36,37] DNA can serve as a cofactor promoting protein folding and DNA-binding function. For example, nonspecific DNA promotes helical structure in MaxbHLH-Fos and MaxbHLH-FosW, but not in MaxbHLH-FosLA, whereas the E-box DNA target promotes folding of all three hybrids. Similarly, sequence-specific binding by bZIP transcription factors induces dimeric and stabilized protein structure,[54,55] and increases stability of the protein structure in many bHLH and bHLHZ proteins.[47,49,56,57] Therefore, the structural and functional attributes displayed by our artificial transcription factors and by native proteins are similarly affected by their environment and assay conditions. All of these extrinsic factors, together with the intrinsic properties of the protein, will influence folding and DNA binding, making the relationship a complicated one between protein structure and DNA-binding function.

Conclusion

In summary, we demonstrated for the first time that a c-Fos LZ domain-containing protein is able to homodimerize under certain conditions. Due to the inability of the c-Fos LZ to homodimerize in its native bZIP form, we speculate that the homodimerization of c-Fos LZ in these bHLHZ-like hybrid proteins is a *forced* homodimerization, in contrast to native dimerization occurring in the Jun/Fos LZ pair and C/EBP LZ pair. By comparing the E-box binding and secondary structures of three MaxbHLH-Fos derivatives that differ only in targeted mutations in the c-Fos LZ, we showed that although proper folding of the four-helix bundle in the HLH domain could force formation of the c-Fos LZ homodimer, the dimerization potency of the LZ region in bHLHZ proteins affects the readiness of overall folding and DNA binding. Therefore, the bHLHZ structure can use its LZ not only to dictate partner specificity,[16,44] but also to fine-tune the ability of the dimeric bHLHZ transcription factor to bind its DNA target, thereby regulating gene expression.

The context-dependent homodimerization of the c-Fos LZ in bHLHZ-like hybrid MaxbHLH-Fos provides a valuable model demonstrating that communication between protein subdomains can yield orthogonal differences in structure and/or function between different protein motifs. In particular, a subdomain may adopt a conformation or activity that is otherwise disfavored when the subdomain is isolated or part of a different protein motif. Context-dependent structure and function also suggests that the space that Nature has explored during evolution is indeed sparse, which leaves much room for design of engineered proteins with tailored functions: such further work depends on clarification of the basic mechanisms for how proteins interact with their DNA targets.

Experimental

Construction of HIS3 and lacZ reporter strains

Reporter strains YM4271[pHISi-1/E-box] and YM4271[pLacZi/Ebox] were constructed as described previously.[22,39] Briefly, four tandem copies of the E-box target sequence (5′-CACGTG) were cloned into the pHISi-1 and pLacZ integrating reporter vectors, respectively, which were then integrated into the yeast genome of *S. cerevisiae* YM4271 (Matchmaker™ One-Hybrid System, Clontech, Mountain View, CA) to construct the two reporter strains containing E-box target sequences upstream of the *HIS3* and *lacZ* reporter genes. 10 mM 3-AT (3-aminotriazole, Bioshop, Burlington, ON) was sufficient to suppress the background due to leaky His3 expression in reporter strain YM4271[pHISi-1/E-box].

Gene construction

DNA oligonucleotides were purchased from Operon Biotechnologies (Huntsville, AL) or Integrated DNA Technologies (Coralville, IA). The gene encoding ArntbHLH (residues 65–130) was constructed using the sequence from human Arnt isoform 3 (NCBI NP_848514) as previously described.[21] Note that the earlier manuscript states the numbering as ArntbHLH

(residues 80–145), but the correct numbering is ArntbHLH (residues 65–130) according to human Arnt isoform 3. Alignment of this ArntbHLH (residues 65–130) construct with human Arnt isoform 1 (NCBI NP_001659) indicated that it is the same as residues 80–145 of isoform 1, except for 4 residues (D80E, M82S, N84A, and F89L) that are located upstream of the DNA-binding region and are not directly involved in DNA binding.[31,58] In comparison with ArntbHLH-Fos, ArntbHLH (residues 65–130) also contains an extra Gln81 in addition to the aforementioned four mutations (Fig. 1). As these residues are located upstream of the DNA-binding region, they are unlikely to affect the DNA binding function of the proteins, and therefore, were not anticipated to affect the comparison of ArntbHLH-Fos with the other two constructs. This expectation was corroborated by the results presented in this study (more information about Arnt isoforms is provided in the ESI†).

In this study, Arnt isoform 1 was utilized, as the Arnt hybrid presented here is part of ongoing research (ref. 59; G.C. and J.A.S, unpublished results) in which Arnt isoform 1 is being utilized. Therefore, the construct ArntbHLH-Fos follows isoform 1 numbering. The sequence encoding ArntbHLH-Fos, a hybrid of human ArntbHLH (residues 82–149) fused to human c-Fos LZ (residues 166–202[60]), was constructed in a two-step manner: the TFos fragment containing the sequence encoding amino acids 143–149 of human Arnt followed by amino acids 166–202 of human c-Fos was assembled by self-priming PCR,[61] and the ArntbHLH sequence encoding amino acids 82–149 of the bHLH domain of human Arnt isoform 1 was amplified from the $Arnt_{82\text{-}464}$ cDNA fragment.[59] The final ArntbHLH-Fos fragment was then PCR-amplified with TFos and ArntbHLH described above as templates (Table S1; more details about gene construction and assembly are provided in the ESI†).

The MaxbHLH sequence encoding residues 22–74 of murine Max protein[14] were constructed as previously described.[39] The sequence encoding MaxbHLH-Fos, a hybrid of murine MaxbHLH (residues 22–74) fused to human c-Fos LZ (residues 166–202[60]), was constructed by PCR similar to that described above for ArntbHLH-Fos.

The DNA sequences described above were inserted into vector pGADT7-2 (equivalent to pGAD424-MCS II in ref. 62), a GAL4 AD plasmid constructed by replacing the MCS of pGAD424 (Matchmaker™ One-Hybrid System, Clontech) with that of pGADT7 (Matchmaker™ Two-Hybrid System, Clontech). Recombinant plasmids were then transformed into *E. coli* strain DH5α, harvested, and sequences confirmed by DNA sequencing (The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON).

Spot titration assay using the *HIS3* reporter

The *HIS3* reporter gene was used to compare the transactivation potency of different GAL4 AD fusion proteins. Yeast transformations were performed using the transformation procedure developed by Dohmen *et al.*[63] Fresh colonies of YM4271[pHISi-1/E-box] transformed with an AD-fusion plasmid were resuspended in sterile water and normalized to OD_{600} = 1. Four ten-fold serial dilutions of fresh colony resuspensions were generated. 15 μL of each dilution (10^{-1}–10^{-4}) for each transformant was spotted on SD/-H/-L plates and SD/-H/-L test plates containing 20 mM 3-AT. The relative strengths of reporter gene activation were estimated by comparing the growth status of different transformants with the same cell densities on test plates.

LacZ reporter assays

The X-gal colony-lift filter assay and *ortho*-nitrophenyl-β-galactoside (ONPG) liquid assay were performed to assess the E-box binding of the hybrid proteins described above. Assay protocols are provided in the Yeast Protocols Handbook from Clontech. In the X-gal colony-lift filter assay, the lysed yeast cells were incubated with X-gal for 2 h before being photographed. Several modifications were made to the Clontech ONPG protocol. Greater reproducibility between replicates was seen when the OD_{600} ranges were narrowed.[62] YPDA cultures for the assay were only inoculated from overnight cultures with OD_{600} 1.0–1.3, and growth was stopped when OD_{600} was 0.60–0.64. After washing with Z buffer, the cells were resuspended in one-fifth culture volume of Z buffer for pGADT7-2, MaxbHLH, and ArntbHLH-Fos, and one-half culture volume of Z buffer for MaxbHLH-Fos. Subsequent steps follow the Clontech procedure with all samples incubated with ONPG for 1 h. ONPG values are given in dimensionless β-galactosidase units, defined as the amount that hydrolyzes 1 μmol ONPG to *o*-nitrophenol and D-galactose per min per cell. Results are presented as mean values ± standard error of the mean (SEM) of three independent experiments, each performed in triplicate.

Protein expression

Bacterial expression and purification of ArntbHLH-Fos, MaxbHLH, MaxbHLH-Fos, MaxbHLH-FosLA, and MaxbHLH-FosW proteins were performed similarly to those of the native MaxbHLHZ described previously.[22] A brief summary of this procedure follows. DNA fragments encoding the five proteins in *E. coli*-preferred codons were assembled and cloned into pET-28A(+) (Novagen, Mississauga, ON; details of gene construction are given in the ESI†). Proteins were expressed in BL21(DE3)pLysS (Stratagene), purified by TALON metal ion affinity chromatography (Clontech) and reversed-phase HPLC (Beckman System Gold, Fullerton, CA; C_{18} column, Vydac), and their identities confirmed by ESI-MS (Waters Micromass ZQ, Model MM1). Protein concentrations were determined by UV/Vis spectrometry (Beckman DU 640; ε_{275} = 1405 M^{-1} cm^{-1} per Tyr). The bacterially expressed ArntbHLH-Fos, MaxbHLH-Fos, and MaxbHLH-FosLA do not contain a C-terminal proline (residue 202 of c-Fos) due to a cloning error. We suspect that this missing Pro does not affect the overall structure of the fusion proteins, because it is outside the LZ region and adjacent to the C-terminal His tag, which is highly basic and unlikely to be involved in LZ dimerization.

Fluorescence anisotropy titrations

Oligonucleotides with 6-carboxyfluorescein (6-FAM) incorporated at the 5′-end were synthesized and HPLC-purified by Operon Biotechnologies. Titrations of the Arnt and Max derivatives were performed with the preferred DNA duplex, in which the core E-box sequence CACGTG is flanked with Arnt-preferred base pairs[64]

 This journal is © The Royal Society of Chemistry 2012

and Max-preferred base pairs,[65] respectively (Fig. 3). Fluorescence was measured on a JY Horiba Fluorolog-3 spectrofluorimeter. The cell (Starna, Atascadero, CA) contained 1 nM DNA duplex in Tris buffer for Arnt derivatives (100 mM Tris, 150 mM NaCl, 1 mM EDTA, 200 mM guanidine–HCl, pH 7.4, 20% glycerol, 0.1 mg mL^{-1} acetylated BSA, 100 mM bp calf thymus DNA)[21] or phosphate buffer for Max derivatives (4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, 150 mM NaCl, 2.7 mM KCl, 1 mM EDTA, 800 mM urea, pH 7.4, 20% glycerol, 0.1 mg mL^{-1} acetylated BSA, 100 μM bp calf thymus DNA)[22,39] in a total volume of 0.3 mL. The Arnt derivatives were found to be better-behaved in Tris buffer than in phosphate buffer, most likely due to formation of well-folded, functional structures and reduced aggregation.[21,22,39]

We observed that our proteins can aggregate at low micromolar concentrations, and hence, significant amounts of denaturant were included in the buffers to maintain solubility critical for gaining reliable data. For our proteins, these concentrations of denaturant were the lowest achievable that promoted protein solubility and still maintained DNA-binding function. For each data point, stock protein was added to the cell to final concentration 0–2 μM protein monomer. The volume change was maintained at $<5\%$ of total volume. The sample was incubated at 4 °C overnight followed by >20 min at room temperature; this temperature-leap tactic was used to minimize protein misfolding and aggregation.[66,67] Titrations were performed at 22.0 ± 0.2 °C.

The polarization values were used to calculate apparent dissociation constants using Kaleidagraph 3.6 (Synergy software) and eqn (1):[36]

$$P = ((P_{bound} - P_{free})[M]/(K_d + [M])) + P_{free} \quad (1)$$

where K_d corresponds to the apparent monomeric dissociation constant, M is the concentration of monomeric protein, P_{free} is the polarization of free DNA, and P_{bound} is the maximum polarization of specifically bound DNA. Protein solubility could not be reliably maintained above 2.0 μM protein monomer, so strong saturation plateaus could not be obtained for all curves.

We emphasize that others have experienced difficulty with solubility of Max derivatives, for example: akin to our MaxbHLHZ domain with a C-terminal His tag, Jung *et al.* expressed a similar Max derivative that contained an additional 15 N-terminal residues preceding the basic region.[68] Unlike our experiments, the authors did not use denaturant; however, they could not maintain protein solubility beyond 20 μg mL^{-1} (approximately 1.5 μM), which precluded their ability to examine their protein's structure by CD. Meier-Andrejszki *et al.* even had difficulty maintaining solubility with their expressed *full-length* Max protein, and this impacted their ability to conduct heat capacity measurements in sufficiently concentrated stock solutions.[69] Likewise, although we also experienced difficulty obtaining data at higher protein concentrations, R values >0.975 were obtained for each of the two independent titrations for all K_d values measured. Dissociation constants are presented as the average K_d value ± SEM.

Circular dichroism spectroscopy

Circular dichroism was performed in phosphate buffer, as Tris buffer is not suitable for our CD measurements due to high background in the far-UV wavelength region.[70] Given these buffer restrictions, CD measurements could not be obtained for the Arnt derivatives under the same conditions as used in fluorescence anisotropy. For CD measurements, 1 mL samples were prepared with 0.5 μM or 2 μM protein monomer in 15.1 mM Na_2HPO_4, 4.9 mM KH_2PO_4, 50 mM NaCl, pH 7.4. In samples with DNA, 2 μM protein and 2 μM DNA duplex were added (*i.e.* 2-fold excess of DNA over functional protein dimer), and DNA duplexes were the same as those used in fluorescence anisotropy (Fig. 3) but without the 6-FAM label. The temperature-leap tactic was used to generate functional proteins for CD measurements: samples, including buffer control without protein, were prepared and incubated overnight at 4 °C, followed by >20 min incubation at room temperature. CD was performed on an Aviv 215 spectrometer with a suprasil, 10 mm path-length cell (Hellma, Plainview, NY) at 22 °C. Spectra were acquired between 180 and 300 nm at 0.2 nm increments and a sampling time of 0.2 s. Each spectrum was the average of two scans with the average buffer control spectrum subtracted. Data obtained were not smoothed. Protein helix content was calculated by the method of Chau and coworkers.[71] Briefly, percentage of helix content was determined by assuming only helical content and using the equation $H = \theta_{222}/[\theta_{H222}{}^{\infty}(1 - k_{222}/n)$, where H is the percent helicity, θ_{222} is the mean residue ellipticity at 222 nM, $\theta_{H222}{}^{\infty}$ is the reference value for a helix of infinite length, k_{222} is a wavelength-dependant constant, and n is the number of amino acids in the protein.

Acknowledgements

We thank Mark Nitz for use of his fluorimeter. We are grateful for funding from NIH (R01 GM069041), NSERC Discovery Grant, Canadian Foundation for Innovation/Ontario Innovation Trust (CFI/OIT), and the University of Toronto.

The authors declare no competing financial interests.

References

1 C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.*, 1992, **61**, 1053–1095.
2 P. Blancafort, D. J. Segal and C. F. Barbas III, *Mol. Pharm.*, 2004, **66**, 1361–1371.
3 C. W. Garvie and C. Wolberger, *Mol. Cell*, 2001, **8**, 937–946.
4 T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell (Cambridge, Mass.)*, 1992, **71**, 1223–1237.
5 J. N. M. Glover and S. C. Harrison, *Nature*, 1995, **373**, 257–261.
6 W. Keller, P. König and T. J. Richmond, *J. Mol. Biol.*, 1995, **254**, 657–667.
7 P. König and T. J. Richmond, *J. Mol. Biol.*, 1993, **233**, 139–154.
8 P. F. Johnson, *Mol. Cell. Biol.*, 1993, **13**, 6919–6930.
9 L. J. Ransone and I. M. Verma, *Annu. Rev. Cell Biol.*, 1990, **6**, 539–557.
10 P. Agre, P. F. Johnson and S. L. McKnight, *Science*, 1989, **246**, 922–926.
11 J. W. Sellers and K. Struhl, *Nature*, 1989, **341**, 74–76.
12 P. C. M. Ma, M. A. Rould, H. Weintraub and C. O. Pabo, *Cell (Cambridge, Mass.)*, 1994, **77**, 451–459.
13 T. Ellenberger, D. Fass, M. Arnaud and S. C. Harrison, *Genes Dev.*, 1994, **8**, 970–980.
14 A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff and S. K. Burley, *Nature*, 1993, **363**, 38–45.
15 P. Brownlie, T. A. Ceska, M. Lamers, C. Romier, G. Stier, H. Teo and D. Suck, *Structure*, 1997, **5**, 509–520.
16 S. K. Nair and S. K. Burley, *Cell (Cambridge, Mass.)*, 2003, **112**, 193–205.

17 P. B. Card, P. J. A. Erbel and K. H. Gardner, *J. Mol. Biol.*, 2005, **353**, 664–677.
18 A. Chapman-Smith and M. L. Whitelaw, *J. Biol. Chem.*, 2006, **281**, 12535–12545.
19 R. J. Kewley, M. L. Whitelaw and A. Chapman-Smith, *Int. J. Biochem. Cell Biol.*, 2004, **36**, 189–204.
20 T. H. Scheuermann, D. R. Tomchick, M. Machius, Y. Guo, R. K. Bruick and K. H. Gardner, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 450–455.
21 H.-K. Chow, J. Xu, S. H. Shahravan, A. T. De Jong, G. Chen and J. A. Shin, *PLoS One*, 2008, **3**, e3514.
22 J. Xu, G. Chen, A. T. De Jong, S. H. Shahravan and J. A. Shin, *J. Am. Chem. Soc.*, 2009, **131**, 7839–7848.
23 H. C. Hurst, *Protein Profile*, 1994, **1**, 123–168.
24 J. R. S. Newman and A. E. Keating, *Science*, 2003, **300**, 2097–2101.
25 C. Vinson, A. Acharya and E. J. Taparowsky, *Biochim. Biophys. Acta, Gene Struct. Expression*, 2006, **1759**, 4–12.
26 T. D. Halazonetis, K. Georgopoulos, M. E. Greenberg and P. Leder, *Cell (Cambridge, Mass.)*, 1988, **55**, 917–924.
27 J. M. Mason, M. A. Schmitz, K. M. Muller and K. M. Arndt, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 8989–8994.
28 P. J. Hurlin and J. Huang, *Semin. Cancer Biol.*, 2006, **16**, 265–274.
29 Y. Kikuchi, S. Ohsawa, J. Mimura, M. Ema, C. Takasaki, K. Sogawa and Y. Fujii-Kuriyama, *J. Biochem. (Tokyo)*, 2003, **134**, 83–90.
30 K. Sogawa, R. Nakano, A. Kobayashi, Y. Kikuchi, N. Ohe, N. Matsushita and Y. Fujii-Kuriyama, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 1936–1940.
31 H. I. Swanson, W. K. Chan and C. A. Bradfield, *J. Biol. Chem.*, 1995, **270**, 26292–26302.
32 H. I. Swanson and J.-H. Yang, *J. Biol. Chem.*, 1996, **271**, 31657–31665.
33 S. G. Bacsi, S. Reisz-Porszasz and O. Hankinson, *Mol. Pharmacol.*, 1995, **47**, 432–438.
34 J. C. Rowlands and J.-Å. Gustafsson, *Crit. Rev. Toxicol.*, 1997, **27**, 109–134.
35 R. H. Wenger, D. P. Stiehl and G. Camenisch, *Sci. STKE*, 2005, **306**, re12.
36 J. L. Huffman, A. Mokashi, H. P. Bachinger and R. G. Brennan, *J. Biol. Chem.*, 2001, **276**, 40537–40544.
37 A. Chapman-Smith, J. K. Lutwyche and M. L. Whitelaw, *J. Biol. Chem.*, 2004, **279**, 5353–5362.
38 K. T. O'Neil, J. D. Shuman, C. Ampe and W. F. DeGrado, *Biochemistry*, 1991, **30**, 9030–9034.
39 J. Xu, A. T. De Jong, G. Chen, H.-K. Chow, C. O. Damaso, A. Schwartz Mittelman and J. A. Shin, *Protein Eng., Des. Sel.*, 2010, **23**, 337–346.
40 K. T. O'Neil and W. F. DeGrado, *Science*, 1990, **250**, 646–651.
41 K. J. Lumb and P. S. Kim, *Biochemistry*, 1995, **34**, 8642–8648.
42 K. M. Campbell, A. J. Sholders and K. J. Lumb, *Biochemistry*, 2002, **41**, 4866–4871.
43 D. Porte, P. Oertel-Buchheit, M. John, M. Granger-Schnarr and M. Schnarr, *Nucleic Acids Res.*, 1997, **25**, 3026–3033.
44 T. Ellenberger, *Curr. Opin. Struct. Biol.*, 1994, **4**, 12–21.
45 J. A. R. Worrall and J. M. Mason, *FEBS J.*, 2011, **278**, 663–672.
46 A. Banerjee, H. Hu and D. J. Goss, *Biochemistry*, 2006, **45**, 2333–2338.
47 M. Horiuchi, Y. Kurihara, M. Katahira, T. Maeda, T. Saito and S. Uesugi, *J. Biochem. (Tokyo)*, 1997, **122**, 711–716.
48 P. Bishop, C. Jones, I. Ghosh and J. Chmielewski, *Int. J. Pept. Protein Res.*, 1995, **46**, 149–154.
49 H. Wendt, R. M. Thomas and T. Ellenberger, *J. Biol. Chem.*, 1998, **273**, 5735–5743.
50 J. M. Mason and K. M. Arndt, *ChemBioChem*, 2004, **5**, 170–176.
51 M. J. Smith, D. C. Charron-Prochownik and E. V. Prochownik, *Mol. Cell. Biol.*, 1990, **10**, 5333–5339.
52 G. J. Kato, W. M. Lee, L. L. Chen and C. V. Dang, *Genes Dev.*, 1992, **6**, 81–92.
53 C. Muhle-Goll, M. Nilges and A. Pastore, *Biochemistry*, 1995, **34**, 13554–13564.
54 J. J. Kohler and A. Schepartz, *Biochemistry*, 2001, **40**, 130–142.
55 M. A. Weiss, T. E. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison and K. Struhl, *Nature*, 1990, **347**, 575–578.
56 R. Fairman, R. K. Beran-Steed and T. M. Handel, *Protein Sci.*, 1997, **6**, 175–184.
57 J. W. Cave, K. Werner and D. E. Wemmer, *Protein Sci.*, 2000, **9**, 2354–2365.
58 L. Q. Dong, Q. Ma and J. P. Whitlock, *J. Biol. Chem.*, 1996, **271**, 7942–7948.
59 G. Chen and J. A. Shin, *Anal. Biochem.*, 2008, **382**, 101–106.
60 F. van Straaten, R. Müller, T. Curran, C. Van Beveren and I. M. Verma, *Proc. Natl. Acad. Sci. U. S. A.*, 1983, **80**, 3183–3187.
61 P. J. Dillon and C. A. Rosen, *BioTechniques*, 1990, **9**, 298–300.
62 G. Chen, L. M. DenBoer and J. A. Shin, *BioTechniques*, 2008, **45**, 295–304.
63 R. J. Dohmen, A. W. M. Strasser, C. B. Höner and C. P. Hollenberg, *Yeast*, 1991, **7**, 691–692.
64 H. I. Swanson and J.-H. Yang, *Nucleic Acids Res.*, 1999, **27**, 3205–3212.
65 D. L. C. Solomon, B. Amati and H. Land, *Nucleic Acids Res.*, 1993, **21**, 5372–5376.
66 G. H. Bird, A. R. Lajmi and J. A. Shin, *Biopolymers*, 2002, **65**, 10–20.
67 Y. Xie and D. B. Wetlaufer, *Protein Sci.*, 1996, **5**, 517–523.
68 K. C. Jung, H. S. Rhee, C. H. Park and C.-H. Yang, *Biochem. Biophys. Res. Commun.*, 2005, **334**, 269–275.
69 L. Meier-Andrejszki, S. Bjelic, J.-F. Naud, P. Lavigne and I. Jelesarov, *Biochemistry*, 2007, **46**, 12427–12440.
70 S. M. Kelly, T. J. Jess and N. C. Price, *Biochim. Biophys. Acta, Proteins Proteomics*, 2005, **1751**, 119–139.
71 Y.-H. Chen, J. T. Yang and K. H. Chau, *Biochemistry*, 1974, **13**, 3350–3359.

 This journal is © The Royal Society of Chemistry 2012