

Reengineering natural design by rational design and *in vivo* library selection: the HLH subdomain in bHLHZ proteins is a unique requirement for DNA-binding function

Jing Xu¹, Antonia T.De Jong¹, Gang Chen¹,
Hui-Kwan Chow¹, Christopher O.Damaso¹,
Adrian Schwartz Mittelman¹ and Jumi A.Shin^{2,3}

¹Department of Chemistry, University of Toronto, Mississauga, ON, Canada L5L 1C6 and ²Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada M5S 3G9

³To whom correspondence should be addressed.
E-mail: jumi.shin@utoronto.ca

Received October 29, 2008; revised October 26, 2009;
accepted December 9, 2009

Edited by Dan Tawfik

To explore the role of the HLH subdomain in bHLHZ proteins, we designed sets of minimalist proteins based on bHLHZ protein Max, bHLH/PAS protein Arnt and bZIP protein C/EBP. In the first, the Max bHLH and C/EBP leucine zipper were fused such that the leucine heptad repeats were not in register; therefore, the protein dimerization interface was disrupted. Max1bHLH-C/EBP showed little ability to activate transcription from the E-box (5'-CACGTG) in the yeast one-hybrid assay, and no E-box binding by quantitative fluorescence anisotropy. Max1bHLH-C/EBP's activity was significantly improved after library selection (three amino acids randomized between HLH and leucine zipper), despite the Max bHLH and C/EBP zipper still being out of register: a representative mutant gave a high nanomolar K_d value for E-box binding. Thus, selection proved to be a powerful tool for salvaging the flawed Max1bHLH-C/EBP, although the out-of-register mutants still did not achieve the strong DNA-binding affinities displayed by their in-register counterparts. ArntbHLH-C/EBP hybrids further demonstrated the importance of maintaining register, as out-of-register mutants showed no E-box-responsive activity, whereas the in-register hybrid showed moderate activity. In another design, we eliminated the HLH altogether and fused the Max basic region to the C/EBP zipper to generate bZIP-like hybrids. Despite numerous designs and selections, these hybrids possessed no E-box-responsive activity. Finally, we tested the importance of the loop sequence in MaxbHLHZ by fluorescence and circular dichroism. In one mutant, the loop was shortened by two residues; in the other, the Lys57:DNA-backbone interaction was abolished by mutation to Gly57. Both showed markedly decreased E-box-binding relative to MaxbHLHZ. Our results suggest that, in contrast to the more rigid bZIP, the HLH is capable of significant conformational adaptation to enable gene-regulatory function and is required for protein dimerization and positioning the basic region for DNA recognition.

Keywords: Arnt/C/EBP/conformational adaptability/
fluorescence anisotropy/Max

Introduction

Increased efforts have focused on development of artificial transcription factors (TFs) to control and manipulate gene expression (Nagaoka and Sugiura, 2000; Ansari and Mapp, 2002; Blancafort *et al.*, 2004). Among TFs, the basic region/leucine zipper (bZIP), basic-region/helix-loop-helix (bHLH) and basic-region/helix-loop-helix/leucine zipper (bHLHZ) families share significant structural similarities: a highly conserved α -helical DNA-binding element (basic region), as well as strongly helical dimerization elements comprising either a leucine zipper (LZ) or HLH subdomain, or both HLH and LZ fused together (Ferre-D'Amare *et al.*, 1993; Ma *et al.*, 1994; Miller *et al.*, 2003). Despite their structural similarities, these three different TF motifs continue to coexist in nature.

As we work toward design of small, minimalist protein structures capable of specific DNA-targeting function and potential regulation of gene expression, we questioned why *both* the HLH and LZ are necessary for structure and function in the bHLHZ family, while the HLH *or* LZ is sufficient for promoting strong dimerization affinity and dictating partner specificity in protein homo- and/or heterodimerization in the bHLH or bZIP families, respectively. Thus, we questioned whether the HLH subdomain from the bHLHZ motif can function independently, as in the bHLH family, and also whether it can be removed altogether to generate a more minimalist bZIP-like structure that retains native DNA-binding function.

We focused on the bHLH subdomains from bHLHZ protein Max and bHLH/PAS (Per-Arnt-Sim) protein Arnt. Max is a highly characterized TF for which high-resolution crystal structures exist (Ferre-D'Amare *et al.*, 1993; Brownlie *et al.*, 1997; Nair and Burley, 2003). The Myc/Max/Mad transcriptional network comprises widely expressed bHLHZ proteins critical for normal cell proliferation and differentiation (Amati and Land, 1994; Orian *et al.*, 2003), and *myc* genes are suspected of being among the most frequently affected in human tumors and disease (Nesbit *et al.*, 1999). Heterodimeric Myc/Max is a transcriptional activator that binds the Enhancer box (E-box, 5'-CACGTG) (Blackwood and Eisenman, 1991; Blackwell *et al.*, 1993). Myc does not homodimerize (Amati and Land, 1994), so its activity is mediated by heterodimerization with Max, which can homodimerize and bind to the E-box site, as well (Blackwood and Eisenman, 1991). Similarly, Arnt heterodimerizes with

various partners including the aryl hydrocarbon receptor and oxygen sensor HIF-1 α , and thereby governs their transcriptional activities (Hoffman *et al.*, 1991; Reyes *et al.*, 1992; Schmid *et al.*, 2004).

Like Max, Arnt can homodimerize and bind to the E-box site; Max and Arnt homodimers may antagonize their normal heterodimeric functional activities in the cell (Swanson and Yang, 1999; Swanson, 2002). Minimalist proteins based on Max and Arnt may provide a protein-based tool for targeting the E-box DNA site, which is involved in 50% or more of known cancers and tumors (Ponzielli *et al.*, 2005). We therefore concentrated on design of small proteins based on the bZIP, bHLH and bHLHZ families to target the E-box. These proteins were achieved by a combination of rational design and *in vivo* library selection in yeast. Such minimalist designs may enable us to elucidate the additional roles of the HLH subdomain aside from protein dimerization, as well as the importance of the LZ as a dimerization element in the context of the bHLHZ structure.

By rational design, we generated bHLHZ-like Max and Arnt derivatives, which comprise the Max or Arnt bHLH fused to the C/EBP LZ. We used *in vivo* yeast one-hybrid assays (Y1H) and *in vitro* circular dichroism (CD) and fluorescence anisotropy to show that maintaining the leucine heptad repeat register between the HLH and LZ, i.e. a seamless hydrophobic dimerization interface, is necessary for strong binding to the E-box DNA site. However, by *in vivo* library selection, we were able to uncover mutants of an out-of-register protein with much improved ability to target the E-box. We also found that the bHLHZ motif cannot be simplified to a bZIP-like structure by removal of the HLH subdomain, even after extensive testing of different hybrid designs and library selections in yeast. Furthermore, two rationally designed loop mutants of MaxbHLHZ point to the key role played by the loop toward protein structure that leads to DNA-binding function, as assayed by CD and fluorescence anisotropy. In addition, these results also indicate that library selection can be useful for discovery of new mutants that would otherwise not be uncovered by rational design alone.

Materials and methods

Detailed experimental information is available in the Supplementary data available at *PEDS* online.

Construction of *HIS3* and *lacZ* reporter strains

Four tandem copies of the E-box target sequence (5'-CACGTG) were cloned into the pHISi-1 integrating reporter vector at the *his3-200* locus of *Saccharomyces cerevisiae* YM4271 (Matchmaker™ One-Hybrid System, Clontech, Mountain View, CA, USA). Ten micromolar 3-AT (3-aminotriazole, Bioshop, Burlington, ON, Canada) was sufficient to suppress background due to leaky His3 expression in reporter strain YM4271[pHISi-1/E-box]. Similarly, reporter strain YM4271[pLacZi/Ebox] was constructed with four copies of the E-box site upstream of the *lacZ* reporter gene.

Construction of genes

DNA oligonucleotides were purchased from Operon Biotechnologies (Huntsville, AL, USA). The genes for expression of native MaxbHLHZ [92 amino acids (aa),

residues 22–113 (Ferre-D'Amare *et al.*, 1993)], native MaxbHLH, MaxbHLH-C/EBP[Met], MaxbHLH-C/EBP[Thr], MaxbHLH-C/EBP[Ile], ArntbHLH-C/EBP[Arg], ArntbHLH-C/EBP[Arg-Ile] and ArntbHLH-C/EBP[Ile-Arg] were synthesized in a single PCR reaction (Wu *et al.*, 2006). Amplified gene inserts were inserted into vector pGAD424 (Matchmaker™ One-Hybrid System). Recombinant plasmids were transformed into *Escherichia coli* strain DH5 α , harvested and correct sequences were confirmed by DNA sequencing.

Generation of the MMbHLH/XXX library

For the MMbHLH/XXX (mutant Max1bHLH-C/EBP) library, in which the 3-aa linker between Helix 2 and the leucine zipper was randomized, the gene encoding the three randomized aa was constructed from two oligonucleotides. One chemically synthesized oligonucleotide containing nine randomized bases (5'-T CAA ATC TTG GAC AAG GCT ACC GAA TAC ATC CAA TAC ATG NNN NNN NNN ACC CAA CAA AAG GTT TTG GA-3') in the linker was annealed to a second oligonucleotide (5'-T GTC GTT GTC AGA GGT CAA TTC CAA AAC CTT TTG TTG GGT-3') at a centrally overlapping 20 bp region and extended in both directions by mutually primed synthesis with DNA Polymerase, Klenow fragment (New England Biolabs, Pickering, ON, Canada). Three hundred nanograms of vector pGAD424/Max1bHLH-C/EBP were linearized at the BamHI site and cotransformed with fragments containing the randomized linker bases. These fragments have 40 bp homology with the 5' and 3' ends of the linearized plasmid.

Theoretically, NNN degeneracy at 3 aa positions would yield a combinatorial library of 8000 protein variants ($20 \times 20 \times 20$) encoded by 2.6×10^5 unique DNA sequences [$(4 \times 4 \times 4)^3$]. The actual library size, according to the colony number on the titer plate, was determined to be 10^5 transformants. Therefore, it is unclear that all 2.6×10^5 unique DNA sequences were covered in this transformation; however, an underrepresented library can also be sufficient for isolation of specific binders, as Nord *et al.* (1995) successfully isolated a number of proteins with proper structure and solubility from an underrepresented library of α -helical bacterial receptor domains. We note that sequences of the selected variants indicate that codon preference in yeast was not a major factor in this selection (Supplementary Table S1 available at *PEDS* online).

Transformation of yeast cells

For the *HIS3* assays, we developed an electroporation protocol based on the methods of Suga and Hatakeyama (Suga and Hatakeyama, 2001, 2003). For the library transformation, 300 ng linearized plasmid plus a 5-fold molar excess of PCR product were transformed. After electroporation, cells were plated on minimal selective medium (SD, synthetic dropout) lacking Leu and His with the appropriate amount of 3-AT to suppress background (Matchmaker™ One-Hybrid System). For assays using the *lacZ* reporter, plasmids were transformed into the integrating reporter strain YM4271[pLacZi/E-box] by the TRAF0 method (Dohmen *et al.*, 1991).

Library selection and validation of positive clones

Positive colonies of at least 2 mm diameter appeared after 4–6 days in the *HIS3* assay. The first validation was to replat positive clones. Plasmids were extracted from the

where K_d corresponds to the apparent monomeric dissociation constant, M is the concentration of monomeric protein, P_{free} the polarization for free DNA and P_{bound} the maximum polarization of specifically bound DNA; two independent titrations (R values >0.975) were performed for each K_d value \pm SEM (standard error of the mean). No binding by any protein was detected with the non-specific DNA control, even at $2 \mu\text{M}$ monomeric protein concentration (data not shown). Despite our use of the temperature-leap tactic, protein solubility was a persistent problem at higher concentrations. Achieving fully saturated equilibrium binding isotherms was difficult, given that the proteins tested showed very weak affinity for the E-box site. Hill coefficient analyses of E-box-binding isotherms showed that MMbHLH/ERY and MaxKG have similar values to that of MaxbHLHZ; this indicates that these proteins, like MaxbHLHZ, likely form dimeric structures for cooperative binding to the E-box site.

Circular dichroism spectroscopy

One milliliter of samples was prepared with $2 \mu\text{M}$ protein monomer in $15.1 \text{ mM Na}_2\text{HPO}_4$, $4.9 \text{ mM KH}_2\text{PO}_4$, 50 mM NaCl , pH 7.4. 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 at least 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, USA) 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 (Chen *et al.*, 1974).

Results

We used both the *HIS3* and *lacZ* reporters for testing all of our designed constructs in the Y1H assay. Activation of the *HIS3* reporter was confirmed by growth on medium lacking histidine, whereas activation of the *lacZ* reporter was detected by two different colorimetric assays: qualitative X-gal colony-lift filter assay and quantitative ONPG liquid assay. Although yeast reporter assays do not provide direct, quantitative detection of the binding of our hybrids to the E-box, the transcriptional readout from reporter activation generally correlates with protein:DNA binding. Therefore, the Y1H provides a satisfactory system for testing the *in vivo* DNA-binding function of our hybrids (Chen *et al.*, 2008; Chow *et al.*, 2008; Xu *et al.*, 2009).

Hybrids of the Max bHLH and C/EBP leucine zipper demonstrate the importance of maintaining the hydrophobic protein dimerization interface

We initially generated the truncated MaxbHLH construct with no leucine zipper. MaxbHLH did not show any growth in the *HIS3* assay or any reporter activation in the *lacZ*-based assays (data not shown), whereas the full MaxbHLHZ DNA-binding domain is capable of strong activation in these assays (Xu *et al.*, 2009). Thus, the LZ is required for Max to achieve the proper structure leading to activation from the E-box. We therefore hypothesized that the HLH must 'cooperate' with the LZ, meaning that the alignment of the

hydrophobic dimerization interface extending from the second helix (Helix 2) in the HLH subdomain through the adjacent LZ must be preserved by maintaining the register of the leucine heptad repeats (leucine/hydrophobic residue every seventh aa, Fig. 1). By preserving the protein dimer's structural integrity, the HLHZ region serves as a stable scaffold for properly orienting the basic regions to achieve the requisite DNA-binding function in the major groove.

We tested our hypothesis by designing three hybrids. The first hybrid Max1bHLH-C/EBP comprises the full MaxbHLHZ domain with its native leucine zipper swapped with the C/EBP LZ; at the junction between the Max bHLH and C/EBP LZ lies the non-native RIR linker, which encodes a BamHI restriction site that facilitates swapping of subdomains. Although LZ swapping between bHLHZ protein Max and bZIP protein C/EBP is a fairly conservative change, the insertion of the non-native RIR linker breaks the periodicity of the Leu heptad repeat by a four-residue displacement. As a result, the hydrophobic dimerization interface extending from Helix 2 in the HLH subdomain through the adjacent LZ is disrupted; the register of the hydrophobic face of Helix 2 is shifted by 51° with respect to that of the C/EBP LZ (Ferre-D'Amare *et al.*, 1993; Brownlie *et al.*, 1997). We anticipated that Max1bHLH-C/EBP would show minimal ability to activate transcription from the E-box in the Y1H.

The second and third hybrids, MaxbHLH-C/EBP[Met] and MaxbHLH-C/EBP[Thr], maintain the correct register of the hydrophobic heptad repeats. They are direct fusions of the Max bHLH subdomain and C/EBP LZ with no linker. The only difference lies in the last aa of Helix 2: methionine in MaxbHLH-C/EBP[Met] originates from the last aa of Helix 2 of MaxbHLHZ, whereas threonine in MaxbHLH-C/EBP[Thr] comes from the C/EBP LZ. A fourth hybrid, MaxbHLH-C/EBP[Ile], is a cloning artifact, in which isoleucine becomes the last aa of Helix 2. Because these hybrids maintain the register of the hydrophobic interface extending from Helix 2 through the LZ, we anticipated that their ability to activate transcription from the E-box would be strong, for none of the three residues (Met, Thr or Ile) is known for being particularly disruptive of α -helical protein structure (O'Neil and DeGrado, 1990).

All four hybrids of the Max bHLH and C/EBP LZ showed weak to strong colony growth under histidine auxotrophy in the *HIS3* assay in the Y1H (Fig. 2), indicating measurable transcriptional activation from the E-box. In the X-gal colony lift assay, the hybrids displayed light to vivid blue color (Fig. 3). As this assay is not quantitative, the ONPG liquid assay was also performed to quantify expression of the *lacZ* reporter from activation at the E-box site. The negative control pGAD424 and positive control native MaxbHLHZ gave ONPG values 7.0 ± 0.5 and 147.4 ± 7.3 , respectively (Fig. 4) (Xu *et al.*, 2009). The first hybrid Max1bHLH-C/EBP gave a very low ONPG reading of 10.3 ± 0.5 as expected, for the non-native RIR linker at the junction of the swap does not maintain the register of the dimerization interface in the native bHLHZ structure. *In vitro* fluorescence anisotropy analysis showed no binding of the E-box site by Max1bHLH-C/EBP (data not shown). In contrast, the aligned MaxbHLH-C/EBP[Met], MaxbHLH-C/EBP[Thr] and MaxbHLH-C/EBP[Ile] showed comparable ONPG values of 95.3 ± 4.0 , 102.6 ± 9.5 and 86.2 ± 5.0 , respectively, all of which are much higher than that of Max1bHLH-C/EBP, but still noticeably lower than that for MaxbHLHZ.

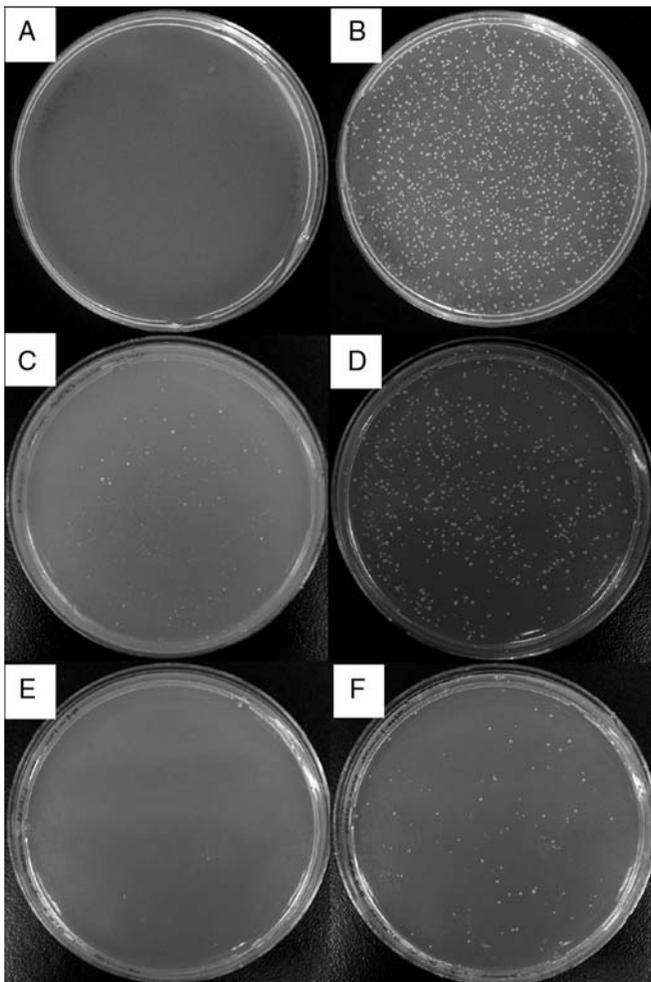


Fig. 2. The *HIS3* assay. All constructs were expressed in pGAD424. SD/-His/-Leu plates were incubated at 30°C, 6 days. (A) pGAD424 (negative control) on 10 mM 3-AT; clean. (B) Native MaxbHLHZ (positive control) on 10 mM 3-AT; strong colony growth. Note the colonies were too crowded to grow larger. Same amounts were plated on all plates for comparison. (C) Max1bHLH-C/EBP on 10 mM 3-AT; weak growth. (D) MaxbHLH-C/EBP[Met] on 10 mM 3-AT; strong growth. MaxbHLH-C/EBP[Thr] and MaxbHLH-C/EBP[Ile] showed similar growth. (E) Max1bHLH-C/EBP on 60 mM 3-AT; clean. (F) MMbHLH/EYR on 60 mM 3-AT; strong growth. The other four selected constructs MMbHLH/ERY, MMbHLH/EWV, MMbHLH/NEV and MMbHLH/NRN also showed comparably strong growth as MMbHLH/EYR. Note that the photos are less clear than the actual plates. (Enlarged version is shown in Supplementary Fig. S4 available at *PEDS* online.)

Arnt bHLH and C/EBP LZ hybrids also confirm the importance of maintaining the register of Leu heptad repeats

We can compare the four hybrids of the Max bHLH and C/EBP LZ described above with our previously designed ArntbHLH-C/EBP, a hybrid comprising the bHLH from bHLH/PAS protein Arnt and leucine zipper from bZIP protein C/EBP (i.e. an interfamily swap of the 29 aa C/EBP LZ replacing the 330 aa Arnt PAS domain) (Chow *et al.*, 2008). ArntbHLH-C/EBP is expected to be bHLHZ-like, although there are no high-resolution structures of the bHLH domain from the bHLH/PAS family. Thus, our design was based on the assumption that the HLH subdomain of Arnt is structurally similar to those of the related bHLH and bHLHZ TFs and is therefore predicted to have a similar mode of DNA binding

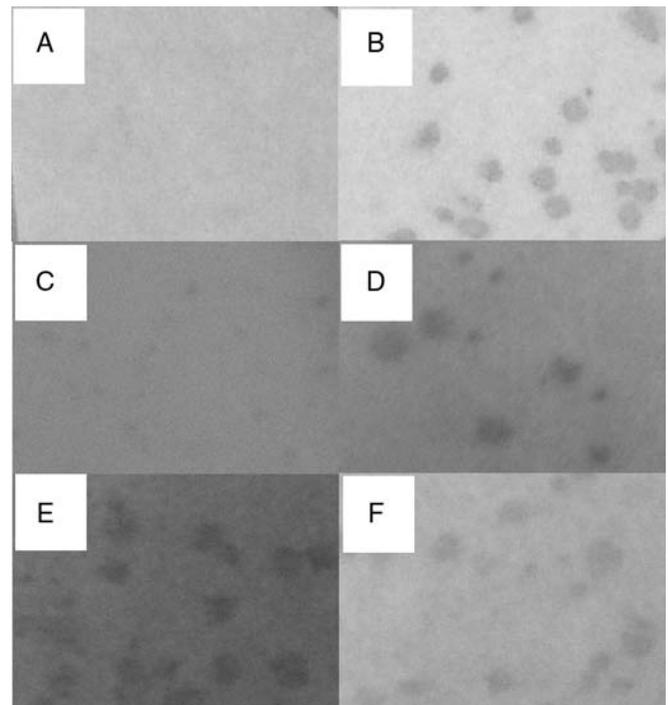


Fig. 3. The X-gal colony-lift filter assay. All SD/-Ura/-Leu plates were incubated at 30°C, 4 days before testing. Photos taken after 2 h incubation. Note that color intensity is affected by variations in colony size and plate density, and blue color intensity in photos is less vivid than actual filter paper. (A) pGAD424 (negative control); very faint blue. (B) Native MaxbHLHZ (positive control); dark blue. (C) Max1bHLH-C/EBP; light blue. (D) MMbHLH/EYR; dark blue. The other four selected constructs MMbHLH/ERY, MMbHLH/EWV, MMbHLH/NEV and MMbHLH/NRN also showed comparable blue color to MMbHLH/EYR. (E) MaxbHLH-C/EBP[Met]; dark blue. (F) MaxbHLH-C/EBP[Thr]; dark blue. (Color version is shown in Supplementary Fig. S3 available at *PEDS* online.)

(Chapman-Smith *et al.*, 2004). ArntbHLH-C/EBP, in which the register of the Leu heptad repeats is maintained between the Arnt bHLH and C/EBP LZ, gives moderate β -galactosidase activity of 44.5 ± 7.4 from activation at the E-box and shows a reasonable free energy of binding to the E-box by fluorescence anisotropy titrations (K_d 148.9 ± 2.9 nM) (Chow *et al.*, 2008).

To further examine the importance of maintaining the register of the hydrophobic dimerization interface at the fusion between the bHLH and LZ, we generated three Arnt hybrids with linkers of variable lengths at the junction between Helix 2 and the LZ. In contrast to the Max derivatives above, the 'in-register' design ArntbHLH-C/EBP possesses the three-residue linker RIR at the Helix2/LZ junction. In comparison, the one- or two-residue linkers R, RI and IR at the junctions of ArntbHLH-C/EBP[Arg], ArntbHLH-C/EBP[Arg-Ile] and ArntbHLH-C/EBP[Ile-Arg] cause these mutants to be out of register. Transcriptional readout from these constructs in the Y1H assays was very weak. No colony growth was observed in the *HIS3* assay, and only very pale blue color was detected in the colony-lift assay (data not shown). The ONPG assay confirms these negative results, as all values were statistically the same as for the negative control pGAD424. Thus, these mutants of the Arnt bHLH and C/EBP LZ also confirmed the importance of maintaining the register of the hydrophobic dimerization interface in the bHLHZ structure, whether in hybrids of the Max bHLH and C/EBP LZ to produce a native-like bHLHZ structure or in hybrids of the Arnt bHLH

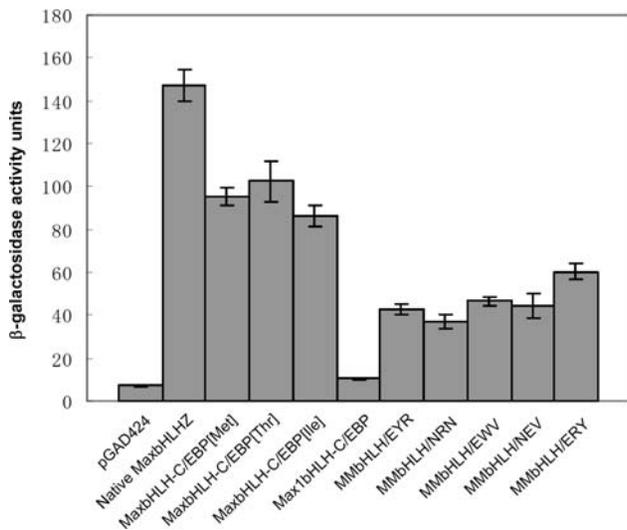


Fig. 4. ONPG histogram comparing transcriptional activation strengths from the E-box site. All values are averages of at least 9–12 measurements (\pm standard deviation) from three to four separate cell-growth cultures.

and C/EBP LZ to generate a ‘non-native’ bHLHZ structure from a bHLH/PAS protein.

Transcriptional activation from the E-box by the hybrid with the disrupted dimerization interface was significantly improved by in vivo selection

The out-of-register Max1bHLH-C/EBP showed weak E-box transcriptional readout in the Y1H assays and no E-box binding *in vitro*. As discussed above, the almost complete loss of function in Max1bHLH-C/EBP was largely caused by a disruption of the Leu/hydrophobic residue heptad repeats in the dimerization region.

However, the relative orientation of the HLH and LZ subdomains may be slightly changed with a three-residue linker of different identity. The RIR linker in Max1bHLH-C/EBP is not particularly effective at maintaining helical structure, for Arg is a relatively weak α -helix former and stabilizer (O’Neil and DeGrado, 1990). Thus, we considered the possibility of uncovering stronger E-box binders by selection from a library of Max1bHLH-C/EBP derivatives with varying aa composition in the linker. We therefore generated library MMbHLH/XXX, with the RIR linker replaced by a randomized pool of 3-aa linkers.

We emphasize that this library still keeps the disrupted hydrophobic heptad repeat structure, for the Helix 2 and LZ are still out of register. We focused on such a library for two reasons. First, because this shift in register is not severe, that is, Helix 2 and the LZ are shifted by only 51° , we could test the ability of the linker to compensate for the flawed, disrupted dimerization domain. Even a short linker can play an essential role in maintaining cooperative inter-domain interactions (Gokhale and Khosla, 2000). An exploitable, convenient property in design of TFs is their modularity, and the correct design of linkers joining these modules is crucial for desired function; library selection may produce better linkers than rational design. Second, we surmised that a good target range to assess would be between the ONPG values of Max1bHLH-C/EBP and the positive MaxbHLHZ control, which is known to be a strong binder of the E-box site (Hu

et al., 2005; Jung *et al.*, 2005; Meier-Andrejszki *et al.*, 2007; Xu *et al.*, 2009). To facilitate our library selections, we developed an improved Y1H system by making use of homologous recombination and developing a rapid and practical electroporation method for transformation of yeast, thereby allowing for fast and straightforward selection of proteins that target a specific DNA site (see Materials and methods and Supplementary data available at *PEDS* online).

After one round of selection, seven positive clones were chosen, and all carried unique sequences. Of these seven sequences, five clones containing 3-aa linkers were selected: MMbHLH/EYR, MMbHLH/ERY, MMbHLH/EWV, MMbHLH/NEV and MMbHLH/NRN (the 3-aa linkers are EYR, ERY, EWV, NEV and NRN, respectively). Interestingly, the other two positive clones contained linkers that also maintained a 51° displacement at the Helix 2/LZ junction and are fully discussed in the Supplementary data available at *PEDS* online. The selected 3-aa mutants showed significantly higher E-box-responsive activity than Max1bHLH-C/EBP. In the *HIS3* assay, Max1bHLH-C/EBP displayed weak colony growth on 10 mM 3-AT (Fig. 2), limited growth on 20 mM 3-AT and no growth on 40 mM 3-AT (data not shown). In contrast, MMbHLH/EYR showed strong growth even on 60 mM 3-AT (Fig. 2). The other four selected constructs also showed similarly strong growth as MMbHLH/EYR.

Both the *lacZ*-based colony-lift and ONPG assays confirmed the *HIS3* assay results. Compared with Max1bHLH-C/EBP (ONPG 10.3 ± 0.5 , light blue color in colony-lift), all five constructs developed bright blue color in the colony-lift assay and ONPG readings of 42.8 ± 2.3 , 60.4 ± 4 , 36.9 ± 3.6 , 46.6 ± 2.2 and 44.6 ± 5.6 , respectively, for MMbHLH/EYR, MMbHLH/ERY, MMbHLH/NRN, MMbHLH/EWV and MMbHLH/NEV (Figs 3 and 4). These values were much improved over Max1bHLH-C/EBP, but not as strong as the native MaxbHLHZ control.

By quantitative fluorescence anisotropy analysis, MMbHLH/ERY, which had the highest ONPG value among the mutants, bound to the E-box with K_d 861 ± 71 nM, an affinity ~ 60 -fold lower than native MaxbHLHZ with K_d 14.3 ± 7.9 nM (Xu *et al.*, 2009) (Fig. 5). Thus, by selection performed on a library purely focused on the 3-aa linker at the junction of the bHLH and LZ, we were able to uncover a few linkers leading to significantly improved activation from the E-box by our out-of-register hybrid. Thus, an improved linker can significantly compensate for flaws in protein structure that affect function.

The HLH subdomain in bHLHZ proteins is a unique requirement for DNA-binding function

To further explore the role of the HLH in bHLHZ proteins, we also investigated whether the HLH can be replaced by a zipper-like structure. We discussed above all our designs that contain the HLH: the truncated MaxbHLH (no LZ), the out-of-register Max1bHLH-C/EBP and selected derivatives (EYR, ERY, NRN, EWV and NEV), and the in-register MaxbHLH-C/EBP derivatives (Met, Thr and Ile). Below, we discuss two different designs in which the HLH has been removed to give a bZIP-like structure of the Max basic region directly fused to either the C/EBP or Max leucine zipper.

We explored two different minimalist protein designs. The first included hybrids between the Max basic region and

Oligonucleotides for fluorescence anisotropy**E-box site, 24-mer**5' - (6-FAM) TGCAGGAACCACGTGGTTGAAGGTT**Non-specific duplex, 24-mer**

5' - (6-FAM) TGCAGGAATTCCAAGGTGAAGGTT

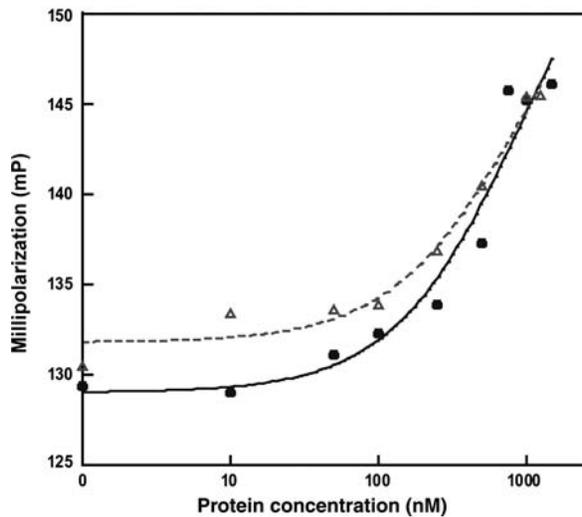


Fig. 5. (Top) DNA duplexes used in fluorescence anisotropy titrations. '6-FAM' is 6-carboxyfluorescein, and the Max-preferred E-box is underlined (core E-box is CACGTG). (Bottom) Representative equilibrium binding isotherms for MMbHLH/ERY (filled circle, solid line) and MaxKG (open triangle, dashed line) targeting the E-box. Each isotherm was obtained from an individual titration, and each K_d value is the average of two individual titrations \pm SEM (see Material and methods).

C/EBP LZ, which were anticipated to yield simple bZIP-like structures (Max1-, Max2- and Max3-C/EBP; Figs 1 and S1, Supplementary data available at *PEDS* online). The second design simplified the MaxbHLHZ to a bZIP-like structure directly by mutations and deletions in the Max HLHZ subdomain (Supplementary Fig. S1 available at *PEDS* online). Our initial hypothesis was that if the long α -helix comprising much of the dimerization interface was maintained, as depicted in the Max:E-box crystal structures in which Helix 2 and the adjacent LZ form a long, seamless α -helix (Ferre-D'Amare *et al.*, 1993; Brownlie *et al.*, 1997) and that is similar to that of the purely α -helical bZIP that dimerizes via a leucine-zipper coiled coil (Ellenberger *et al.*, 1992; König and Richmond, 1993; Glover and Harrison, 1995), our minimalist hybrids should be capable of reporter gene activation from the E-box in the Y1H assay.

In both studies, we performed rational design and extensive *in vivo* library selection (see Supplementary Fig. S1 available at *PEDS* online which gives a complete listing of sequences and libraries tested, and supplemental results from selections are discussed). For the first design, we focused on properly orienting the DNA-binding basic regions that project from the LZ into the DNA major groove by inserting 1, 2 or 3 aa between the dimerization and DNA-binding elements. An α -helix contains 3.6 residues per turn, so up to 3 aa should reasonably cover the circular orientation possibilities in the major groove: hence, Max1-C/EBP, Max2-C/EBP and Max3-C/EBP that differ by 1 aa at the C-terminal end of the Max basic region. For the second design, mutations and deletions in the HLHZ subdomain of MaxbHLHZ were designed to maintain the Leu heptad

repeat structure throughout the LZ (Supplementary Fig. S1 available at *PEDS* online). We also incorporated mutations in the Max LZ that were demonstrated to increase the dimerization strength of native Max (Tchan and Weiss, 2001).

No transcriptional readout from the E-box could be detected by either design. We used two approaches in an attempt to improve the first design, including random mutagenesis on the entire protein gene and focusing the library on the hinge region between the basic region and leucine zipper, as the hinge was found to be important for orienting the basic region in the major groove (see Supplementary Fig. S1 available at *PEDS* online) (Pu and Struhl, 1991). For the second design, library selections focused on optimizing the linker between the Max basic region and the LZ, which had been simplified from the HLHZ (see Supplementary Figs S1 and S2 and Table S2 available at *PEDS* online). However, even after extensive testing and selection, we could not detect by transcriptional readout a clear, unambiguous interaction between the E-box and any of the bZIP-like designs.

Using Miller *et al.*'s (2003) alignment of bZIP proteins for which crystal structures have been determined, we observed that the spacing between the invariant Arg in the basic region and the first aa of the leucine zipper is 9 aa in length. For our designed Max-C/EBP hybrids, this spacing is between 12 and 14 aa, which is longer by approximately one turn of the α -helix. Because these residues are from MaxbHLHZ Helix 1, they might not be a suitable linker for the bZIP-like structure. To rule out this possibility, we instead used 9 aa from C/EBP to generate Max-C/EBP-C/EBP, which shares high sequence similarity to C/EBP with the conserved basic residues at the same positions (Fig. S1A and B available at *PEDS* online). However, Max-C/EBP-C/EBP did not give higher transcriptional readout in the *HIS3* and *lacZ* assays. This result suggests that this 9-aa spacer is not the reason for observing no DNA-binding function from the bZIP-like Max-C/EBP derivatives.

Circular dichroism and fluorescence anisotropy demonstrated that the loop is important for DNA binding

We speculated that the HLH subdomain in bHLHZ proteins positions each basic region optimally in the DNA major groove; in particular, the loop plays a critical role. The loops in the bHLHZ family are 5–23 residues in length (Ferre-D'Amare *et al.*, 1993). Given such differences in length and the lack of sequence conservation in the loops—in contrast to the highly conserved sequences of Helices 1 and 2 flanking the loops—we did not anticipate the loop playing a critical role in achieving proper structure leading to DNA-binding function. In order to test the importance of the loop in MaxbHLHZ, we constructed two mutants: MaxKG, in which the Lys57 contact to the DNA backbone was abolished by mutation to Gly57, and Max Δ AS, in which Ala58 and Ser59, which are the last two loop residues, were deleted to give a 6-aa loop.

We assessed the intrinsic helical structure present in the MaxbHLHZ loop mutants by CD. MaxKG was much more helical (48%) than Max Δ AS (23%) and even more helical than native MaxbHLHZ (39%, Fig. 6). Such helicity is indicative of a more well-folded, helical structure, but interestingly, the E-box-binding affinity of MaxKG is 697 ± 192 nM, which is greatly decreased compared with native MaxbHLHZ (Fig. 5). The CD results indicated that the K57G

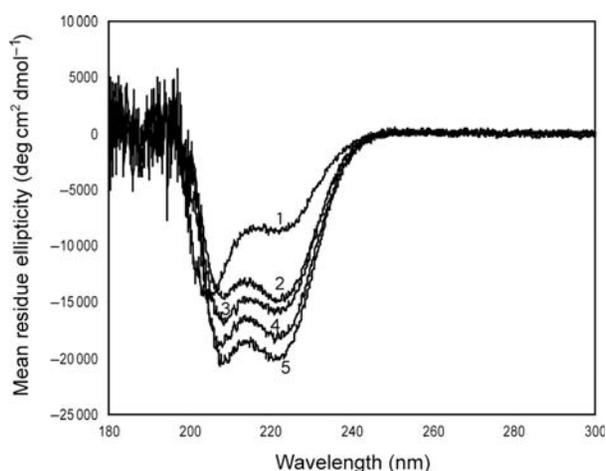


Fig. 6. CD spectra (from top to bottom) of (1) Max Δ AS, (2) MaxbHLHZ, (3) Max1bHLH-C/EBP, (4) MaxKG and (5) MMbHLH/ERY. Each spectrum was averaged twice, and curves were not subjected to smoothing. The buffer control was subtracted from each protein spectrum. Mean residue ellipticities are presented, which account for differences in lengths of the proteins.

loop mutation had no negative effect on helicity and that loop flexibility in MaxKG was probably maintained and comparable to that of MaxbHLHZ, such that formation of the four-helix bundle in the HLH subdomain was properly facilitated. Therefore, structure was likely maintained in MaxKG, but E-box binding function was not; the 50-fold drop in E-box binding affinity by MaxKG should therefore be largely attributable to loss of the Lys57 interaction with the DNA phosphodiester backbone. In contrast, the loop deletions in Max Δ AS did negatively affect the secondary structure of the protein, which is likely due to less flexibility in the truncated loop detrimentally affecting formation of the four-helix bundle in the HLH. Max Δ AS exhibited no binding to the E-box even at 2 μ M monomer (data not shown), likely due to loss of structure capable of DNA-binding function.

The two out-of-register hybrids MMbHLH/ERY and Max1bHLH-C/EBP (with the RIR linker) displayed good helical structure (52% and 42%, respectively). We hypothesized that because the Leu heptad repeats are not in register in Max1bHLH-C/EBP and thus the hydrophobic interface is disrupted, then the ability of this hybrid protein to fold properly and adopt stable helical structure in the dimerization interface would be compromised. CD demonstrated that MMbHLH/ERY, which was selected from a small library of mutants, is noticeably more helical than the template Max1bHLH-C/EBP. This correlates with the observed increase in E-box-binding measured by fluorescence anisotropy.

Discussion

The HLHZ dimerization domain persists in nature despite existence of simpler motifs

The bHLH protein family, including the E proteins and MyoD (Ellenberger *et al.*, 1994; Ma *et al.*, 1994), is able to homo- or heterodimerize and effectively target specific DNA sites without secondary dimerization elements. However, in the bHLHZ family, both the HLH and leucine zipper are required for effective protein dimerization and concomitant DNA-binding function. The question we posed is whether

the HLH itself from bHLHZ proteins is sufficient for DNA binding, and how critical is stringent maintenance of the HLHZ structure in bHLHZ proteins for effective protein dimerization and DNA binding. The truncated MaxbHLH construct, with no LZ, displayed no activation from the E-box. This result suggests that the Max HLH subdomain must cooperate with an adjacent LZ for efficient dimerization and DNA binding.

Since the isolated HLH in bHLHZ proteins cannot act as an independent dimerization domain, why does it still persist in nature, especially given that the bZIP, with just the LZ dimerization element, and the bHLH, with just the HLH dimerization subdomain, are effective transcriptional regulators possessing strong dimerized structure and DNA-binding function? All these protein motifs use basic α -helices to bind the DNA major groove and share significant structural similarity in comparison with other TF families. Recently, we showed that the basic region of Max can be fused directly to the HLH subdomain of bHLH protein E47. This Max-E47 hybrid retains E-box-binding activity similar to the native MaxbHLHZ as measured *in vitro* and *in vivo* (ONPG values and low nM K_d values are virtually identical) (Xu *et al.*, 2009).

Thus, Max's E-box binding function can also be served by a bHLH structure, rather than the native bHLHZ motif, suggesting that the same basic region can function comparably well when fused to either the HLH or HLHZ dimerization subdomain. This result also indicates that although the HLH subdomains in the native MaxbHLHZ and Max-E47 hybrid appear similar in overall structure, both the HLH and HLHZ dimerization elements may coexist due to differences in fine structure that lead to subtle differences in function.

Why is the more complicated HLHZ structure utilized for dimerization? Unless the HLH subdomain in bHLHZ proteins plays other roles, the far simpler leucine zipper should ably serve as a dimerization element. We suspect that the role of the HLH subdomain differs between the bHLH and bHLHZ families. In the bHLH family, the HLH serves multiple purposes: it is responsible both for protein dimerization and for positioning the basic region in the DNA major groove, although how the HLH discriminates among different dimerization partners is unclear (Fairman *et al.*, 1997). But in bHLHZ proteins, the HLH and LZ may have separate roles. The primary role of the HLH in bHLHZ proteins may be to position each basic region in the optimal orientation for interaction with its DNA half site: Ellenberger and coworkers addressed that 'this positioning effect could be the basis for the observed cooperativity of binding to DNA' (Wendt *et al.*, 1998). The role of the LZ is to specify its dimerization partner, whether homo- or heterodimer (Muhle-Goll *et al.*, 1995; Nair and Burley, 2003). Specifically, the LZ can allow better regulation of dimerization specificity, for it can contain aa that are suboptimal for stability but favor interaction with a particular partner (Pabo and Sauer, 1992). Thus, nature may continue to use the HLHZ structure, in comparison with the simpler HLH, for fine-tuning dimerization partner specificity and affinity.

The identity of the leucine zipper is important for DNA-binding activity

Given that the difference in ONPG values between native MaxbHLHZ and the MaxbHLH-C/EBP[Met], [Thr] and [Ile]

derivatives was quite large at ~ 1.5 -fold, we considered it unlikely that the unoptimized Met, Thr or Ile residue between the HLH and LZ elements could be the dominant source of the weaker activation ability. Our analysis indicated that the LZ is the important factor. One might consider that the Max LZ is 39 aa whereas the C/EBP LZ is only 29 aa, and thus, the shorter dimerization interface presented by the C/EBP LZ could lead to a less stable dimer. However, this seems unlikely, because the C/EBP LZ is known to be a strong coiled-coil dimerization element, with a dimerization constant of $17 \mu\text{M}$ (O'Neil *et al.*, 1991); in comparison, the Max LZ, without the HLH subdomain, exhibits a dimerization constant of $\sim 200 \mu\text{M}$ (Muhle-Goll *et al.*, 1995). So this explanation would not explain why replacement of the Max LZ with the C/EBP LZ is significantly detrimental to DNA-binding activity.

We therefore considered that the Max LZ, having coevolved with the HLH, may be optimized particularly for the bHLHZ structure: both the HLH and LZ subdomains cooperate to achieve dimerized structure, and neither subdomain alone is capable of proper dimerization. In comparison, the C/EBP LZ originates from the bZIP motif in which the LZ is directly fused to the DNA-binding basic region: the C/EBP LZ would be optimal for the bZIP structure that depends solely on its leucine zipper for dimerization. Additionally, in MaxbHLHZ, the LZ has coevolved with the flexible HLH to achieve overall dimerized structure, in contrast to the bZIP, with its more rigid, purely α -helical structure.

Maintaining the register of the hydrophobic dimerization interface is advantageous, but the HLH can partly compensate for misalignment

A proper hydrophobic dimerization interface is important for DNA-binding function, as binding to the E-box by the out-of-register Max1bHLH-C/EBP was not measurable by fluorescence anisotropy, and β -galactosidase activity was barely detected in the Y1H. Interestingly, the comparably high ONPG values and sequence similarities of MaxbHLH-C/EBP[Met], MaxbHLH-C/EBP[Thr] and MaxbHLH-C/EBP[Ile] suggest that maintaining proper protein structure by preserving the register of the hydrophobic dimerization interface, not the identities of individual aa, mainly determines the ability for transcriptional activation.

However, when the HLH and LZ are not in proper register to maintain the hydrophobic dimerization interface, DNA binding may still result from the ability of the HLH to adapt its structure to optimize protein:protein and protein:DNA interactions, as observed in the selected hybrids MMbHLH/EYR, MMbHLH/ERY, MMbHLH/EWV, MMbHLH/NEV and MMbHLH/NRN. In particular, the loop and packing of the tetrameric bundle of α -helices in the HLH can afford conformational adaptability to the overall bHLHZ structure. In comparison, the LZ is a more rigid structure; this fact may explain the necessity of a precisely optimal alignment of the basic region and leucine zipper when designing bZIP-like TFs, and that even with extensive rational design and *in vivo* selections on different libraries, we could not salvage the bZIP-like designs. We also note that the α -helical bZIP structure is a highly conserved TF structural motif across species, and the bZIP:DNA crystal structures show that the protein backbone structures of various bZIP proteins are

virtually superimposable (Ellenberger *et al.*, 1992; König and Richmond, 1993; Glover and Harrison, 1995), suggesting less flexibility in the purely helical bZIP structure.

Our results suggest that even when the hydrophobic dimerization interface is no longer in register, the HLH retains the conformational adaptability necessary for DNA-binding function to still occur, that is, it is capable of optimizing protein structure necessary for gene-regulatory function. Despite the absence of high-resolution structures of our protein:DNA complexes, we believe the positioning effect of the HLH to be a highly important one.

Despite lack of conservation, the loop appears to have significant effect on DNA binding

The 'positioning effect' is a unique property of the HLH; in particular, the loop appears to play the critical role in establishing structure that positions the basic regions properly for DNA binding. Both bHLH and bHLHZ proteins show strong conservation of sequences and structures in the basic regions, Helices 1 and 2, and leucine zippers. In contrast, the loops display disordered structures with unconserved lengths and sequences (Ferre-D'Amare *et al.*, 1993; Ma *et al.*, 1994; Brownlie *et al.*, 1997; Nair and Burley, 2003). Such properties may lead us to dismiss the importance of the loop structure, especially given that the loop is not well resolved in any crystal structures, and therefore, its contributions to function are difficult to assess. However, the opposite may be the case, as the loop's flexibility and location enable the protein to adjust its structure at the critical junction between the dimerization scaffold and DNA-binding regions.

Other studies of linkers in various protein families have concluded that linkers lack regular secondary structure and display varying degrees of flexibility to match their particular biological purpose (George and Heringa, 2002), and the loop can serve well the role of flexible linker. We found such conformational adaptability with the selected derivatives of Max1bHLH-C/EBP. In contrast, we were not successful in the design or selection of bZIP-like hybrids capable of activation from the E-box site (see Supplementary data available at *PEDS* online).

CD and fluorescence anisotropy experiments shed light on the relationship between protein structure and DNA-binding function. Although helical secondary structure was largely maintained in MMbHLH/ERY and Max1bHLH-C/EBP, their E-box-binding affinities were noticeably lower than native MaxbHLHZ. We suspect that although reasonably helical in structure, their HLHZ dimerization subdomains did not optimally position the basic regions in the DNA major groove.

Our structure–function analysis also demonstrated that although helical secondary structure in the MaxKG mutant was maintained, its E-box-binding affinity was decreased by 50-fold likely due to loss of the Coulombic interaction between Lys57 in the loop and the DNA backbone. The MaxbHLHZ:E-box structure was the first protein:DNA crystal structure elucidated for a bHLH superfamily protein, and it showed this loop:DNA contact that Burley and co-workers speculated contributed to overall DNA binding (Ferre-D'Amare *et al.*, 1993). Interestingly, no other bHLH superfamily proteins show any loop contacts to DNA. We emphasize that with all our designs, the Max basic region was kept intact, and it should be sufficient for strong and specific binding of the E-box, as shown in the

high-resolution structures (Ferre-D'Amare *et al.*, 1993; Brownlie *et al.*, 1997; Nair and Burley, 2003) and by our Max-E47 hybrid, which targets the E-box as effectively as does native MaxbHLHZ but lacks any loop interactions with DNA. However, in the context of the MaxbHLHZ structure, the Lys57:DNA backbone interaction cannot be disrupted without a large decrease in binding affinity. In comparison, Max Δ AS showed that loop truncation affects both protein structure and DNA-binding function. We suspect that the shortened loop lacks the flexibility to allow proper folding of the HLH structure and cannot position Lys57 properly with respect to the DNA backbone.

In conclusion, our study suggests that although maintaining the register of the hydrophobic dimerization interface between Helix 2 and the LZ is key for creating an effective DNA-binding protein with bHLHZ-like structure, the out-of-register hybrids that were selected in the *in vivo* Y1H can also target the E-box effectively, indicating significant conformational adaptability in the bHLHZ structure. Our results also indicate that the bHLHZ cannot be simplified to a bZIP-like structure. Instead, our study provides additional evidence that the HLH subdomain in bHLHZ proteins has its own unique role: to position each basic region optimally within the DNA major groove.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

The authors thank Alevtina Pavlenco for technical assistance, and Prof. Mark Nitz for use of the spectrofluorimeter.

Funding

This work was supported by National Institutes of Health (R01 GM069041); Premier's Research Excellence Award (PREA); Canadian Foundation for Innovation/Ontario Innovation Trust (CFI/OIT); NSERC Discovery Grant and University of Toronto.

References

- Amati,B. and Land,H. (1994) *Curr. Opin. Genet. Dev.*, **4**, 102–108.
 Ansari,A.Z. and Mapp,A.K. (2002) *Curr. Opin. Chem. Biol.*, **6**, 765–772.
 Bird,G.H., Lajmi,A.R. and Shin,J.A. (2002) *Biopolymers*, **65**, 10–20.
 Blackwell,T.K., Huang,J., Ma,A., Kretzner,L., Alt,F.W., Eisenman,R.N. and Weintraub,H. (1993) *Mol. Cell. Biol.*, **13**, 5216–5224.
 Blackwood,E.M. and Eisenman,R.N. (1991) *Science*, **251**, 1211–1217.
 Blancafort,P., Segal,D.J. and Barbas,C.F., III (2004) *Mol. Pharm.*, **66**, 1361–1371.
 Brownlie,P., Ceska,T.A., Lamers,M., Romier,C., Stier,G., Teo,H. and Suck,D. (1997) *Structure*, **5**, 509–520.
 Chapman-Smith,A., Lutwyche,J.K. and Whitelaw,M.L. (2004) *J. Biol. Chem.*, **279**, 5353–5362.
 Chen,Y.-H., Yang,J.T. and Chau,K.H. (1974) *Biochemistry*, **13**, 3350–3359.
 Chen,G., DenBoer,L.M. and Shin,J.A. (2008) *Biotechniques*, **45**, 295–304.
 Chow,H.-K., Xu,J., Shahravan,S.H., De Jong,A.T., Chen,G. and Shin,J.A. (2008) *PLoS ONE*, **3**, e3514.
 Dohmen,R.J., Strasser,A.W.M., Höner,C.B. and Hollenberg,C.P. (1991) *Yeast*, **7**, 691–692.
 Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) *Cell*, **71**, 1223–1237.
 Ellenberger,T., Fass,D., Arnaud,M. and Harrison,S.C. (1994) *Genes Dev.*, **8**, 970–980.
 Fairman,R., Beran-Steed,R.K. and Handel,T.M. (1997) *Protein Sci.*, **6**, 175–184.
 Ferre-D'Amare,A.R., Prendergast,G.C., Ziff,E.B. and Burley,S.K. (1993) *Nature*, **363**, 38–45.
 George,R.A. and Heringa,J. (2002) *Protein Eng.*, **15**, 871–879.
 Glover,J.N.M. and Harrison,S.C. (1995) *Nature*, **373**, 257–261.
 Gokhale,R.S. and Khosla,C. (2000) *Curr. Opin. Chem. Biol.*, **4**, 22–27.
 Hoffman,E.C., Reyes,H., Chu,F., Sander,F., Conley,L.H., Brooks,B.A. and Hankinson,O. (1991) *Science*, **252**, 954.
 Hu,J., Banerjee,A. and Goss,D.J. (2005) *Biochemistry*, **44**, 11855–11863.
 Huffman,J.L., Mokashi,A., Bachinger,H.P. and Brennan,R.G. (2001) *J. Biol. Chem.*, **276**, 40537–40544.
 Jung,K.C., Rhee,H.S., Park,C.H. and Yang,C.-H. (2005) *Biochem. Biophys. Res. Commun.*, **334**, 269–275.
 König,P. and Richmond,T.J. (1993) *J. Mol. Biol.*, **233**, 139–154.
 Lajmi,A.R., Lovrencic,M.E., Wallace,T.R., Thomlinson,R.R. and Shin,J.A. (2000a) *J. Am. Chem. Soc.*, **122**, 5638–5639.
 Lajmi,A.R., Wallace,T.R. and Shin,J.A. (2000b) *Protein Expr. Purif.*, **18**, 394–403.
 Ma,P.C.M., Rould,M.A., Weintraub,H. and Pabo,C.O. (1994) *Cell*, **77**, 451–459.
 Meier-Andrejszki,L., Bjelic,S., Naud,J.-F., Lavigne,P. and Jelesarov,I. (2007) *Biochemistry*, **46**, 12427–12440.
 Miller,M., Shuman,J.D., Sebastian,T., Dauter,Z. and Johnson,P.F. (2003) *J. Biol. Chem.*, **278**, 15178–15184.
 Muhle-Goll,C., Nilges,M. and Pastore,A. (1995) *Biochemistry*, **34**, 13554–13564.
 Nagaoka,M. and Sugiura,Y. (2000) *J. Inorg. Biochem.*, **82**, 57–63.
 Nair,S.K. and Burley,S.K. (2003) *Cell*, **112**, 193–205.
 Nesbit,C.D., Tersak,J.M. and Prochownik,E.V. (1999) *Oncogene*, **18**, 3004–3016.
 Nord,K., Nilsson,J., Nilsson,B., Uhlén,M. and Nygren,P.A. (1995) *Protein Eng.*, **8**, 601–608.
 O'Neil,K.T. and DeGrado,W.F. (1990) *Science*, **250**, 646–651.
 O'Neil,K.T., Shuman,J.D., Ampe,C. and DeGrado,W.F. (1991) *Biochemistry*, **30**, 9030–9034.
 Orian,A., *et al.* (2003) *Genes Dev.*, **17**, 1101–1114.
 Pabo,C.O. and Sauer,R.T. (1992) *Annu. Rev. Biochem.*, **61**, 1053–1095.
 Ponzielli,R., Katz,S., Baryte-Lovejoy,D. and Penn,L.Z. (2005) *Eur. J. Cancer*, **41**, 2485–2501.
 Prendergast,G.C., Lawe,D. and Ziff,E.B. (1991) *Cell*, **65**, 395–407.
 Pu,W.T. and Struhl,K. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 6901–6905.
 Reyes,H., Reisz-Porszasz,S. and Hankinson,O. (1992) *Science*, **256**, 1193–1195.
 Schmid,T., Zhou,J. and Brüne,B. (2004) *J. Cell. Mol. Med.*, **8**, 423–431.
 Suga,M. and Hatakeyama,T. (2003) *Curr. Genet.*, **43**, 206–211.
 Suga,M. and Hatakeyama,T. (2001) *Yeast*, **18**, 1015–1021.
 Swanson,H.I. (2002) *Chem. Biol. Interact.*, **141**, 63–76.
 Swanson,H.I. and Yang,J.-H. (1999) *Nucleic Acids Res.*, **27**, 3205–3212.
 Tchan,M.C. and Weiss,A.S. (2001) *FEBS Lett.*, **509**, 177–180.
 Wendt,H., Thomas,R.M. and Ellenberger,T. (1998) *J. Biol. Chem.*, **273**, 5735–5743.
 Wu,G., Wolf,J.B., Ibrahim,A.F., Vadasz,S., Gunasinghe,M. and Freeland,S.J. (2006) *J. Biotechnol.*, **124**, 496–503.
 Xu,J., Chen,G., De Jong,A.T., Shahravan,S.H. and Shin,J.A. (2009) *J. Am. Chem. Soc.*, **131**, 7839–7848.