Molecular BioSystems

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c6mb00795c

The DNA target determines the dimerization partner selected by bHLHZ-like hybrid proteins AhRJun and ArntFos†

Ichiro Inamoto, Gang Chen‡ and Jumi A. Shin*

The molecular basis of protein–partner selection and DNA binding of the basic helix–loop–helix (bHLH) and basic region-leucine zipper (bZIP) superfamilies of dimeric transcription factors is fundamental toward understanding gene regulation. Because these families share structural similarities, we swapped the bHLH and leucine zipper (LZ) modules between families to uncover how individual modules influence protein-partnering and protein:DNA complexation. We previously described ArntFos, a bHLHZ-like hybrid of the bHLH domain from the bHLH/PAS protein Arnt and LZ from the bZIP protein c-Fos, binding to the Arnt E-box site (TCACGTGA) as a homodimer. Herein, we describe a heterodimer between ArntFos and AhRJun, a hybrid of the bHLH domain from AhR and LZ of JunD. We designed AhRJun and ArntFos to heterodimerize, given the strong interaction between native AhR/Arnt and Jun/Fos, but the hybrids showed no preference for hetero- or homo-dimerization in Y2H assays. However, adding a specific DNA target drove the formation of a single dimeric protein species over others. EMSA showed that the AhRJun/ArntFos heterodimer binds to the cognate DNA site XRE1 (TTGCGTG) with K_d = 337 nM. Unexpectedly, the palindromic Arnt E-box drove the binding of the AhRJun/ArntFos heterodimer (K_d = 276 nM)—not the ArntFos homodimer—that binds to the Arnt E-box. However, the dimerization preference switched to the ArntFos homodimer when the variant Max E-box (CCACGTGG) was used. We conclude that the DNA sites themselves are the primary determinants of dimerization specificity for AhRJun and ArntFos, not the JunD and c-Fos LZs, a result that sheds light on the dynamics of protein/protein and protein:DNA interactions and the structural modularity of bHLH and bZIP proteins.

Received 23rd November 2016,
Accepted 10th January 2017

DOI: 10.1039/c6mb00795c

www.rsc.org/molecularbiosystems

Introduction

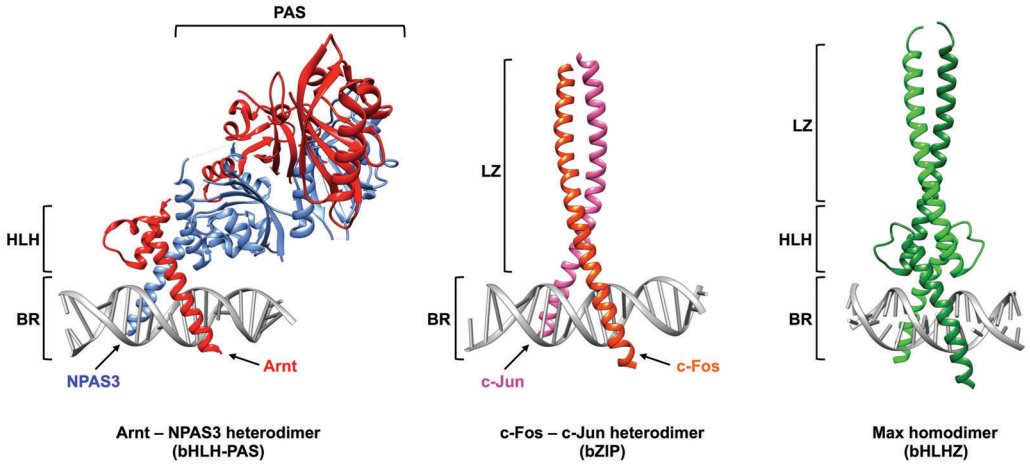
The basic region helix–loop–helix (bHLH) proteins are a large and diverse family of transcription factors that target specific DNA sites for the regulation of genes involved in fundamental processes, such as cell cycle and developmental regulation, apoptosis, and stress response pathways.[1–4] The bHLH proteins typically associate as homo- or hetero-dimers that recognize the hexameric E-box DNA site (enhancer box, CANNTG).[2,5] Members of the bHLH superfamily share a highly conserved structural motif comprising a DNA-binding basic region and a HLH dimerization domain (Fig. 1). The N-terminally located basic region targets the DNA major groove; the C-terminal HLH dimerization domain comprises two amphipathic helices separated by a loop region, typically 5–12 residues in length, which dimerize to form a compact, hydrophobic four-helix bundle positioning the contiguous basic regions for DNA binding.[3,6–8] Two subfamilies of bHLH proteins, bHLH/PAS and bHLHZ, contain additional Per–Arnt–Sim (PAS) and leucine zipper (LZ) domains located C-terminally to the bHLH that further regulates protein dimerization.[9,10]

The aryl hydrocarbon receptor (AhR, also known as the dioxin receptor) and the aryl hydrocarbon receptor nuclear translocator (Arnt) are bHLH/PAS proteins that heterodimerize in the presence of environmental contaminants, such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) and polychlorinated biphenyls (PCBs).[11–15] The AhR/Arnt heterodimer binds to the heptameric xenobiotic response element (XRE, also known as dioxin response element, DRE), TNGCGTG, and activates genes encoding xenobiotic metabolizing enzymes.[11–14,16,17] Biochemical studies have shown AhR to target the TNGC half site with preference for either T or C in the second position.[18,19] Arnt targets the GTG half-site, which incidentally is also the E-box half-site.[18,20] AhR and Arnt have distinctively

Department of Chemistry, University of Toronto, 3359 Mississauga Road, Mississauga, Ontario, Canada L5L 1C6. E-mail: jumi.shin@utoronto.ca; Fax: +1 (905) 828-5425; Tel: +1 (905) 828-5355

† Electronic supplementary information (ESI) available: Plasmid construction details, table of primers, MY1H and MY2H, table of all *in vivo* yeast data, EMSA data and binding isotherms. See DOI: 10.1039/c6mb00795c

‡ 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.




```
                basic region          helix 1       loop        helix 2                 leucine zipper
AhRJun          QKTVKPIPAEGIKSNPSKRHR-DRLNTELDRLASLLP-FPQDVINKLD-KLSVLRLSVSYLRAK-SFFDVALEEKVKTLKSQNTELASTASLLREQVAQLKQKVLSH
AhR(ΔL)Jun      QKTVKPIPAEGIKSNPSKRHR-DRLNTELDRLASLLP-FPQDVINKLD-KLSVLRLSVSYLRAK-SFFDVA-EEKVKTLKSQNTELASTASLLREQVAQLKQKVLSH
ArntFos         SSADKERLARENHSEIERRRR-NKMTAYITELSDMVP-TCSALARKPD-KLTILRMAVSHMKSL-RGTGNTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAH

XRE1            TGCAGGAATTGCGTGATGAAGGTT
Arnt E-box      TGCAGGAATCACGTGATGAAGGTT
Max E-box       TGCAGGAACCACGTGGTGAAGGTT

C/EBP           TGCAGGAATTGCGCAATGAAGGTT
NS DNA          TGCAGGAATTCCAAGGTGAAGGTT
```

Fig. 1 Protein and DNA sequences. (top) Schematic showing structures of the bHLH/PAS Arnt/NPAS1 heterodimer bound to HRE DNA (PDB: 5sy7),[73] the bZIP Jun/Fos heterodimer bound to AP-1 DNA (PDB: 1fos),[74] and the bHLHZ Max homodimer bound to E-box DNA (PDB: 5eyo).[75] BR = basic region, HLH = helix–loop–helix, PAS = Per–Arnt–Sim, LZ = leucine zipper. Images were produced using Chimera, (version 1.11.2).[76] (middle) Sequence alignment of hybrid proteins. The bHLH domains of AhR or Arnt were fused to the leucine zippers of JunD or c-Fos, respectively. The AhR basic region proline that may interact with a thymine in DNA half-site recognition of the Arnt E-box **T**CAC (see Discussion) is highlighted in blue underline. Leu residues that define the LZ motif are in red. The hydrophobic heptad repeat register of helix 2 and LZ is not maintained in AhR(ΔL)Jun, as the first Leu in the Jun LZ is deleted (denoted as "–). Complete details of protein expression are provided in the ESI.† (bottom) DNA target sequences. All oligonucleotides are a double stranded, 24 bp DNA containing a centrally located target site (underlined). Only the forward strand is shown. The core E-box sequence (CAC·GTG, black circle marks the two abutting half-sites) and GTG half-site are shown in red for XRE1, Arnt E-box, and Max E-box.

different preferences for protein dimerization. Arnt dimerizes with many other bHLH/PAS proteins such as AhR, hypoxia inducible factors (HIFs) and single-minded (SIM) proteins, regulating gene expression in their respective physiological pathways.[3] In addition, the Arnt homodimer binds to the canonical E-box site (CACGTG) with preference for sites flanked by T and A (Arnt E-box, TCACGTGA).[20,21] In contrast, AhR, like HIF and SIM, exclusively dimerizes with Arnt and does not homodimerize.[9] The AhR PAS domain plays a central role in the selection of dimerization partners. The AhR bHLH domain (AhRbHLH) expressed without the PAS domain lacked dimerization specificity for Arnt and was capable of dimerizing with Arnt and AhR bHLH/PAS domains. Removal of the PAS domain did not affect DNA binding, as the AhRbHLH/Arnt heterodimer was still bound to the XRE site.[22]

The structures of the bZIP and bHLH families are related:[8,23] instead of dimerization through the HLH structure, however, the bZIP uses the α-helical leucine zipper (LZ) as its dimerization module. The LZ is characterized by a heptad repeat containing Leu at every seventh amino acid. When two LZs dimerize in a parallel coiled coil, the heptad repeat aligns hydrophobic residues on the interface between the monomers, forming a hydrophobic core.[24,25] Charged residues are typically found flanking the hydrophobic core, where a basic residue on one monomer can complement an acidic residue on its partner. These van der Waals and Coulombic interactions are the basis of partner selection and dimerization for bZIP proteins. For example, members of the Fos family of bZIP transcription factors do not homodimerize due to repulsion between the negatively charged Fos LZs.[26] Instead, Fos proteins heterodimerize with members of the Jun family of bZIP proteins to bind to the AP-1 site TGACTCA.[27,28] On the other hand, Jun proteins are capable of weak homodimerization, although heterodimerization with Fos is preferred.[26]

The modularity of individual subdomains in bHLH and bZIP proteins has been demonstrated previously, both within and between protein families. Modularity within the bZIP family has been shown through domain-swap experiments between Jun, Fos, CREB (cyclic AMP response element-binding protein), and GCN4.[29,30] These experiments yielded hybrid proteins that maintained a native α-helical structure and DNA-binding function. Modularity within the bHLH superfamily—comprising bHLH, bHLHZ, and bHLH/PAS protein families—has been reported, but with varying degrees of success.[9,31–33] Previously we described Max-E47,[32] a hybrid protein comprising the Max basic region (Max belongs to the bHLHZ family) and E47 HLH (E47 belongs to the bHLH family), which was shown to be as potent as its

parental proteins in DNA binding affinity and specificity and maintained an α-helical protein structure.[34]

More recently, we generated the hybrid protein ArntFos by fusing the bHLH domain of Arnt to the c-Fos LZ: hence, a bHLHZ-like protein generated from parental proteins belonging to the bHLH and bZIP families.[35] *In vivo* yeast hybrid assays and *in vitro* fluorescence anisotropy showed that ArntFos was capable of homodimerizing and binding to the E-box with DNA sequence specificity, demonstrating successful domain-swapping between the bHLH and bZIP *super*families. ArntFos displayed relatively weak affinity towards the E-box (K_d = 436 nM) that was attributed to electrostatic repulsion between the c-Fos LZs, which do not natively homodimerize. However, the c-Fos LZ was essential for ArntFos activity *in vivo*, as the Arnt bHLH domain alone had no helical structure and was incapable of binding to the E-box in the yeast one-hybrid assay.[31,35]

Thus far, we have focused on hybrid protein homodimers. In this study, we describe a hybrid protein *heterodimer* by heterodimerizing ArntFos with AhRJun, a new bHLHZ-like hybrid generated by fusing the bHLH domain of AhR to the Jun LZ. Since AhR is the native heterodimerization partner of Arnt, and Jun the partner of Fos, AhRJun was expected to form a stable heterodimer with ArntFos. We used *in vivo* yeast one- and two-hybrid (Y1H and Y2H) assays and *in vitro* quantitative electrophoretic mobility shift assays (EMSA) to demonstrate that the AhRJun/ArntFos heterodimer bound to the nonpalindromic XRE target site. However, the protein/protein interactions of AhRJun and ArntFos were promiscuous, with homo- and heterodimers being observed. We confirmed that incubating the hybrids with different DNA target sites resolved this promiscuity, driving exclusive formation of a single dimeric protein:DNA complex over other possibilities. This highlights the strong influence of DNA toward directing the formation of specific protein:protein and protein:DNA complexes for these hybrid proteins.

Experimental

Protein nomenclature

ArntbHLH is just the bHLH domain of human Arnt protein.[36] AhRbHLH is just the bHLH domain of human AhR protein. ArntFos (previously named "ArntbHLH-Fos") is ArntbHLH expressed with a C-terminally fused human c-Fos LZ.[35] AhRJun is AhRbHLH expressed with a C-terminally fused human JunD LZ. AhR(ΔL)Jun is a mutant AhRJun with Leu deletion at the interface of AhRbHLH and JunD LZ (Fig. 1).

Reagents

Reagents were purchased from BioShop Canada (Burlington, ON), enzymes were purchased from New England Biolabs (Pickering, ON), and oligonucleotides were synthesized by Operon Biotechnologies (Huntsville, AL) or Integrated DNA Technologies (Coralville, IA), unless otherwise noted.

Bacterial and yeast strains

Bacterial strains were propagated in LB broth at 37 °C in a shaking incubator at 250 rpm. Ampicillin (50 μg mL^{-1}) and kanamycin (30 μg mL^{-1}) were added as appropriate. The LB agar plates contained 15 g L^{-1} agar. The yeast strains were propagated in either YPD broth or the appropriate minimal synthetic dropout (SD) broth at 30 °C in a shaking incubator at 200 rpm. The agar plates contained 20 g L^{-1} agar.

Escherichia coli DH5α (Stratagene, La Jolla, CA), SURE® (Stratagene) and C2925 (New England Biolabs) were used for standard cloning procedures. The SURE strain was used to handle DNA sequences with unstable secondary structures. The C2925 *dam*$^-$/*dcm*$^-$ strain was used to propagate plasmids free of *dam* and *dcm* methylation to allow cloning using *dam*- or *dcm*-sensitive restriction sites. *E. coli* BL21(DE3)pLysS or Rosetta BL21 was used for the bacterial expression of His-tagged proteins.

Saccharomyces cerevisiae YM4271 [MATa, ura3-52, his3-200, ade2-101, lys2-801, leu2-3, 112, trp1-901, tyr1-501, gal4-Δ512, gal80-Δ538, ade5::hisG] was used as the host for constructing reporter strains for the M1YH assay, as reported previously.[36] A DNA cassette containing six tandem copies of the consensus xenobiotic response element (XRE1, TTGCGTG) placed upstream of the HIS3 or lacZ reporter genes was inserted into the YM4271 genome by recombination, yielding YM4271[pHISi-1/XRE-6] and YM4271[pLacZi/XRE-6], respectively. The 3-aminotriazole (3-AT) titration assay revealed that 30 mM 3-AT was sufficient to inhibit background HIS3 activity in YM4271[pHISi-1/XRE-6] on SD/-His medium. *S. cerevisiae* AH109 [MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::$GAL1_{UAS}$–$GAL1_{TATA}$–HIS3, $GAL2_{UAS}$–$GAL2_{TATA}$–ADE2, URA3::$MEL1_{UAS}$–$MEL1_{TATA}$–lacZ, MEL1] was used for assessing protein/protein interactions in the Y2H system. Both YM4271 and AH109 were purchased from Clontech (Palo Alto, CA).

Transformation, DNA preparation, and plasmid rescue

Standard molecular cloning procedures were performed as described previously.[36,37] PCR reactions were performed with Phusion™ high-fidelity DNA polymerase (New England Biolabs). PCR products and DNA fragments for cloning were purified using the Qiagen DNA purification and Qiagen gel purification kits (Qiagen, Mississauga, ON). For *E. coli*, plasmids were transformed by chemical transformation (TSS protocol, see ref. 38), while yeast transformations were performed using either the standard lithium acetate method (Yeast Protocols Handbook, Clontech) or the transformation procedure developed by Dohmen *et al.*[39] Plasmids were purified using the Wizard® *Plus* SV Minipreps DNA purification system (Promega, Madison, WI) or QIAprep spin miniprep kit (Qiagen, Mississauga, ON).

Plasmid construction

Detailed information on plasmid construction is provided in the ESI.† All new constructs were confirmed by dideoxynucleotide DNA sequencing performed at The Centre for Applied Genomics, The Hospital for Sick Children (Toronto, ON).

Plasmids for modified yeast-one hybrid (MY1H) assays. To identify a protein:DNA interaction, proteins were co-expressed from the GAL4 activation domain (AD) fusion vectors pCETT and pCETT2.[37] Both pCETT and pCETT2 contain two multiple cloning sites (MCSs) where transcription of the inserted genes is under the control of the truncated, low-expression ADH1 promoter.[40]

In pCETT, the gene inserted in MCS I, but not in MCS II, is expressed as the GAL4 AD fusion protein. In pCETT2, proteins in both MCSs are expressed as GAL4 AD fusions. The following nomenclature is used to describe pCETT derivatives: a construct carrying genes in both MCSs is named "pCETT/MCS I/MCS II," a construct carrying a gene in only MCS I is named "pCETT/MCS I," and a construct carrying a gene in only MCS II is named "pCETT//MCS II." Similar nomenclature is used for pCETT2 derivatives.

Plasmids for yeast-two hybrid (Y2H) assays. To identify a protein/protein interaction, proteins were expressed from GAL4 DBD (DNA binding domain) and GAL4 AD fusion vectors, pGBKT7 and pGADT7, respectively (Matchmaker Two-hybrid System 3, Clontech).

HIS3 reporter assay

The yeast strain YM4271[pHISi-1/XRE-6] was transformed with an AD-fusion plasmid. The transformants were plated on minimal synthetic dropout (SD) medium plates lacking His and Leu. To measure the activation of the *HIS3* reporter gene, a fresh colony was resuspended at $OD_{600} \sim 0.01$ in SD media, and 15 μL of the suspension was spotted on SD/-H/-L control plates and SD/-H/-L test plates containing 30 mM 3-AT.

A spot titration assay was used to compare the transactivation potency of different GAL4-AD fusion proteins. Ten-fold serial dilutions of fresh colony resuspensions were generated, where the OD_{600} for each dilution ranged from 0.1 to 0.0001. 15 μL of each dilution was spotted on SD/-H/-L plates and SD/-H/-L test plates containing 30 mM 3-AT. The relative strengths of reporter gene activation were estimated by comparing the growth of different transformants with the same cell densities on the test plates. The growth rate of a transformant in the presence of 3-AT generally correlates with transcriptional activity of the *HIS3* reporter gene.[41]

LacZ reporter assay

The yeast strain YM4271[pLacZi/XRE-6] was transformed with an AD-fusion plasmid. The quantitative *ortho*-nitrophenyl-galactoside (ONPG) liquid assay was performed as described before to quantify the activation of the *LacZ* reporter gene in the transformed cells.[37] Values are expressed in Miller Units of β-galactosidase activity, where 1 unit hydrolyzes 1 μmol ONPG per min per cell. The results are presented as mean values ± standard error of the mean (SEM) of 3–4 independent experiments, each performed in triplicate.

Y2H system

The Matchmaker GAL4 Two-hybrid System 3 (Clontech) was utilized to examine homo- and hetero-dimerization of the proteins under investigation. The recombinant plasmids of the GAL4 DBD fusion vector pGBKT7 and the GAL4 AD fusion vector pGADT7 (or their respective parental vectors in the controls) were co-transformed into the yeast strain AH109. Transformed cells were selected on SD/-L/-W plates at 30 °C, for 4 days. Interactions between the two proteins were determined by simultaneous activation of the *HIS3*, *ADE2*, and *MEL1* reporter genes. Activation of these reporter genes was measured by spotting 15 μL of diluted fresh cell suspensions at $OD_{600} \sim 0.01$ (fresh colonies were resuspended in sterile water and then diluted in 10-fold steps to $OD_{600} \sim 0.01$) on SD/-L/-W plates and on SD/-A/-H/-L/-W/X-α-Gal plates. A positive interaction was indicated by the growth of cultures on both types of plates, with blue colony color showing on the latter plates. The X-α-Gal indicator plates were prepared as per the Yeast Protocol Handbook, Clontech. To exclude false positive signals generated from the direct interaction between our proteins and GAL4 DBD or AD, the yeast strain AH109 was co-transformed with either the pGADT7-based GAL4 AD fusion plasmids and empty pGBKT7, or the pGBKT7-based GAL4 DBD fusion plasmids and empty pGADT7. Co-transformation of empty pGBKT7 and pGADT7 was also tested. Also, in order to exclude the possibility that some proteins under investigation are transcriptionally active, a reverse test was performed that exchanges the two proteins that are fused to DBD and AD.[42]

Protein expression

ArntFos and AhRJun proteins were bacterially expressed and purified following previously described methods, with modifications described below.[32] DNA fragments coding the proteins in *E. coli*-optimized codons were assembled and cloned into the pET-28A(+) expression vector (Novagen; details of gene construction are given in the ESI†). These vectors were transformed into *E. coli* BL21(DE3)pLysS or Rosetta(DE3)pLysS for the expression of AhRJun and ArntFos, respectively. Typically, cells were grown in a 1 L culture where protein production was induced by adding 1 mM IPTG during the mid-log phase of growth ($OD_{600} \sim 0.60$). After induction, the cells were harvested, sonicated, and purified using cobalt metal affinity chromatography (TALON, Clontech) following the manufacturer's protocol. The recommended buffers for TALON typically contain up to 5 mM β-mercaptoethanol (BME). However, BME was excluded from buffers used for ArntFos purification, as mass spectrometry revealed that BME was routinely found covalently attached to a cysteine in ArntFos. After TALON, proteins in the elution fraction were reduced by exposure to 10 mM DTT for 1 h, at 37 °C, and further purified by reversed-phase HPLC (Beckman System Gold, C_{18} column, Vydac). Protein identities were confirmed by ESI-MS (Waters Micromass ZQ, Model MM1), and their concentrations were measured by UV/Vis spectrometry (Nanodrop 2000 Spectrophotometer, Thermo Scientific; ε_{275} = 1405 M^{-1} cm^{-1} per Tyr). Finally, the proteins were lyophilized in aliquots and stored at −80 °C. Immediately before conducting EMSA proteins were reconstituted to the desired concentration (typically 25 or 50 μM) in buffer and incubated for 1 h, at 37 °C, to maximize solubility. Expression and purification of AhR(ΔL)Jun were attempted as well, but the protein could not be expressed in enough quantity for downstream experiments due to its extremely low yield (data not shown). Surprisingly, this was not due to protein insolubility or instability: the single amino acid change resulted in a sharp decrease in protein production.

Electrophoretic mobility shift assay

Single-stranded oligonucleotides containing single copies of the desired protein-recognition site—Arnt or Max E-box,

XRE1, C/EBP, or NS (nonspecific sequence)—were synthesized with 6-carboxyfluorescein (6-FAM) incorporated at their 5′ ends (Operon Biotechnologies, Fig. 1). The 6-FAM-labelled oligonucleotides were annealed to their corresponding unlabeled complementary oligonucleotides by mixing the two in 10 mM Tris–HCl, pH 7, with the unlabeled oligonucleotide in 1.5 molar excess, heating at 95 °C, 10 min, and slow-cooling to room temperature. These annealed DNA targets were used in the following binding reactions.

Protein:DNA binding reactions were performed in EMSA buffer (20 mM Tris–HCl, pH 8; 0.5 mM EDTA; 0.5 mM KCl; 2 mM DTT; 20% glycerol; 100 μg mL^{-1} BSA; and 2 μg mL^{-1} poly dI–dC) and 2 ng of 6-FAM-labeled double-stranded DNA in a total volume of 30 μL. Appropriate amounts of protein solution were added to each reaction to cover the full titration range of the binding reaction. Prior to electrophoresis, samples were treated with the temperature-leap tactic to minimize protein misfolding and aggregation:[43,44] protein solutions were incubated at 4 °C overnight followed by 30 min at 37 °C, and 1 h at room temperature. The samples were loaded onto a pre-equilibrated native PAGE gel (10% poly-acrylamide, 0.5 × TBE), and run at 200 V for 5 min followed by 100 V for 25 min. The gels were visualized using the BioRad ChemiDoc MP Imaging System. Imagelab software (Version 5.2) was used for densitometric analysis for calculation of the bound DNA fraction (θ_{app}) for each lane, where the bound DNA fraction is the intensity of the band corresponding to the protein-bound DNA divided by the sum of the intensities of the bands corresponding to the protein-bound DNA and free DNA. The bound DNA fractions were fit to eqn (1) using KaleidaGraph software (Version 4.5) to calculate the apparent K_d value, as described previously.[45]

$$\theta_{app} = \theta_{min} + (\theta_{max} - \theta_{min})[1/(1 + K_d^2/[M]^2)] \quad (1)$$

All datasets fit to eqn (1) had R values >0.99. Each K_d value was determined from the average of two independent experiments. Hill coefficient analyses showed that all binding isotherms indicate dimeric, cooperative binding of DNA targets with Hill coefficients of approximately 2 (Fig. S5, ESI†).

Results and discussion

Our aim was to generate a pair of bHLHZ-like hybrid proteins that target a specific cognate DNA element as a *heterodimer*. Asymmetric DNA sites are more suited for the investigation of heterodimer:DNA interactions, as these sites are normally not targeted by protein homodimers. The asymmetric XRE site, TNGCGTG, became our ideal candidate. We previously described the bHLHZ-like hybrid protein ArntFos (originally termed "ArntbHLH-Fos") that binds to the E-box site as a homodimer.[35] In this study, we generated another bHLHZ-like hybrid, AhRJun, by combining the AhR bHLH domain (AhRbHLH) and LZ of the JunD protein (Fig. 1). Based on the properties of native AhR/Arnt and Jun/Fos transcription factors, we hypothesized that AhRJun would preferentially heterodimerize with ArntFos, and that the AhRJun/ArntFos heterodimer would bind to the XRE site. We chose the JunD LZ in order to encourage heterodimerization, because JunD has a weaker affinity for homodimerization compared to other Jun family members.[46]

We used the mouse XRE1 sequence, TTGCGTG, since XRE1 was shown to be specifically bound by the AhR/Arnt heterodimer (Fig. 1).[47–49] Similarly, we used the Arnt E-box sequence, TCACGTGA, for investigating ArntFos homodimers, since native Arnt homodimer prefers this E-box sequence.[18,20] AhR alone is incapable of DNA binding. However, the DNA half-site preferred by AhR, T(T/C)GC, suggests that a homodimer of AhR, represented by the AhRJun homodimer, would bind to the C/EBP sequence, TTGCGCAA,[50] prompting us to test this possibility as well.

Co-expression of AhrJun and ArntFos is necessary to activate the XRE1-controlled reporter gene *in vivo*

We first tested whether AhRJun and ArntFos are capable of interacting with the XRE1 site. *HIS3* and *LacZ* reporter assays were utilized to examine our protein design in the modified yeast one-hybrid (MY1H) system that allows co-expression of two proteins that can bind to a DNA target.[37] In MY1H, either one or both of the co-expressed proteins is fused to the GAL4 activation domain (AD). These *in vivo* assays provide an indirect measurement of the DNA-binding affinities of our proteins, because transcriptional potency of reporter activation generally correlates with the strength of the interaction under investigation.[51]

The *HIS3* assay was performed using the yeast strain YM4271[pHISi-1/XRE-6] that possesses the chromosomally integrated reporter cassette in which six tandem copies of XRE1 lie upstream of the *HIS3* reporter gene, allowing *HIS3* expression when the XRE1 site(s) is bound by transcription factors. Reporter gene activation allows cell growth on His-deficient media. Co-expression of GAL4 AD-fused AhRJun (AD-AhRJun) and ArntFos allowed growth on SD/-H/-L plates at 30 mM 3-AT (Fig. S1, ESI†), indicating strong reporter gene activation. In contrast, cells expressing either AD-AhrJun or AD-ArntFos alone were incapable of growth. Co-expression of AD-AhRbHLH and ArntbHLH—which lack the Jun and Fos LZs—also resulted in no growth (Fig. S2, ESI†). Therefore, the presence of both AhRJun and ArntFos was necessary to activate the XRE1-controlled reporter gene in a LZ-dependent manner.

These findings were confirmed by the quantitative *LacZ*-based ONPG assay that uses β-galactosidase activity as an indicator for reporter gene activation (Fig. 2). Proteins were co-expressed in the yeast strain YM4271[pLacZi-1/XRE-6], where six tandem copies of XRE1 lie upstream of the *LacZ* reporter gene. Cells co-expressing AD-AhrJun and AD-ArntFos gave 95.7 ± 3.1 units of β-galactosidase activity, which was significantly higher than the signal produced by the empty vector negative control (unpaired t-test, $P < 0.0001$). Cells expressing either AD-AhRJun or AD-ArntFos alone produced a signal identical to that of the negative control (5–7 units of β-galactosidase activity). Expression of the bHLH domains alone—without the LZ—also did not produce a signal above that of the negative control (data not shown). Again, reporter gene activation was observed only when both AhRJun and ArntFos were present. This suggests that AhRJun and ArntFos were bound to the XRE1 target as a heterodimer.

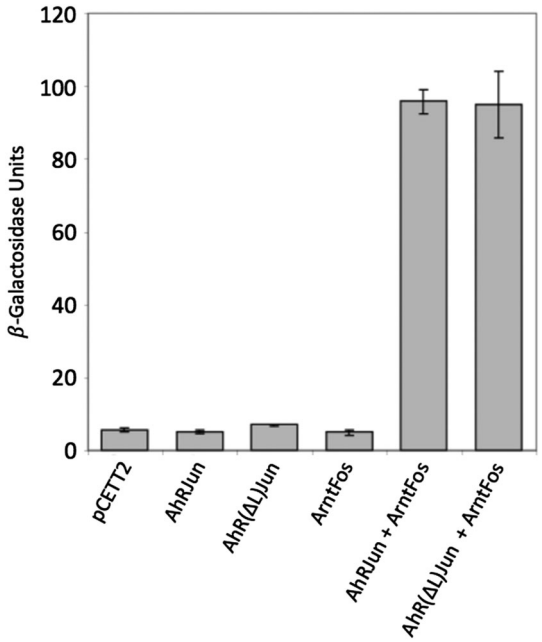


Fig. 2 Both AhRJun and ArntFos are necessary to activate the XRE1-controlled reporter gene in MY1H. The solution containing the 1 : 1 AhRJun : ArntFos interaction with XRE1 was measured in the MY1H assay. Protein expression vector pCETT2 carrying AhRJun, AhR(ΔL)Jun, and ArntFos in different combinations was transformed into the yeast strain YM4271[pLacZi-1/XRE-6]. Error bars represent standard error of the mean from at least three independent trials conducted in triplicate. All bars above contain the vector pCETT2; the first bar is the negative control containing only pCETT2 with no protein expression.

AhRJun and ArntFos show no preference for hetero- or homo-dimer formation *in vivo*

To test whether AhRJun and ArntFos could heterodimerize, we examined their protein/protein interactions using the Y2H system. AhRJun and ArntFos fused to either the GAL4 AD or GAL4 DNA binding domain (DBD) were co-expressed in various combinations in the yeast strain AH109. Interactions between AD- and DBD-fused proteins result in reporter gene activation, allowing colony formation and blue color development on selective media.

Cells co-expressing various combinations of proteins were spot-plated in order to screen for reporter gene activation (Fig. S4 and Table S2, ESI†). First, cells co-expressing DBD-AhRJun and AD-ArntFos showed growth and color development, indicating reporter gene activation (Fig. S4, ESI† Row A). Swapping the GAL4 domains between these proteins, *i.e.* co-expression of AD-AhRJun and DBD-ArntFos, still resulted in reporter activation. In contrast, expression of AhRJun or ArntFos alone did not result in colony growth, regardless of which GAL4 domains were fused to the protein. The negative control co-expressing AD and DBD by themselves did not produce any signal as well (Fig. S4, ESI† row D4). Therefore, it appears that AhRJun and ArntFos heterodimerize in the yeast model system. We then tested if these proteins were capable of forming homodimers (Fig. S4, ESI† rows C5 and C6). Both co-expression of DBD-AhRJun and AD-AhRJun, and DBD-ArntFos and AD-ArntFos, resulted in reporter gene activation, indicating that AhRJun and ArntFos homodimerize *in vivo*. To discern the relative strengths of reporter gene activation, cells co-expressing AD-AhRJun and DBD-ArntFos, AD-AhRJun and DBD-AhRJun, or AD-ArntFos and DBD-ArntFos, were tested by plating cells in serial dilution (Fig. 3). Contrary to our expectation, cells expressing either the AhRJun or ArntFos homodimers produced a signal that was similar to that of cells co-expressing AhRJun and ArntFos. Thus, AhRJun and ArntFos were both similarly capable of hetero- and homo-dimerization in a yeast model.

We find it intriguing that AhRJun and ArntFos showed little dimerization preference. The ability of the c-Fos and JunD LZs to direct specific protein/protein interactions appears to be compromised in these hybrid proteins. This loss of specificity could be caused partly due to the fact that the JunD and c-Fos LZ modules evolved in the context of bZIP protein structure, not as a fusion to bHLH as in our hybrids. Loss of specificity could also be caused by the lack of PAS domains that, together with HLH, forms the dimerization interface between native AhR and Arnt. Even with its PAS domain, Arnt bHLH/PAS is known for its promiscuous protein/protein interactions, forming stable heterodimers with other bHLH/PAS proteins such as AhR, HIF and SIM,[3,52] as well as homodimers.[18,21] The AhR bHLH/PAS, on the other hand, is more stringent in partner specificity,[22] but this is largely dependent on its PAS domain that is missing in our proteins. This promiscuity in protein–partner preference from AhRbHLH and ArntbHLH could be offsetting any selectivity conferred by the JunD and c-Fos LZs. These results present an interesting contrast with the MY1H data above, where the AhRJun/ArntFos heterodimer was the only species capable of binding to the XRE1 site.

The AhRJun/ArntFos heterodimer binds to XRE1 with high affinity and specificity *in vitro*

We next used quantitative EMSA to explore the interaction of AhRJun and ArntFos against XRE1 (Table 1). For solutions containing both AhRJun and ArntFos, proteins were mixed in 1 : 1 stoichiometry: for example, a 600 nM AhRJun/ArntFos

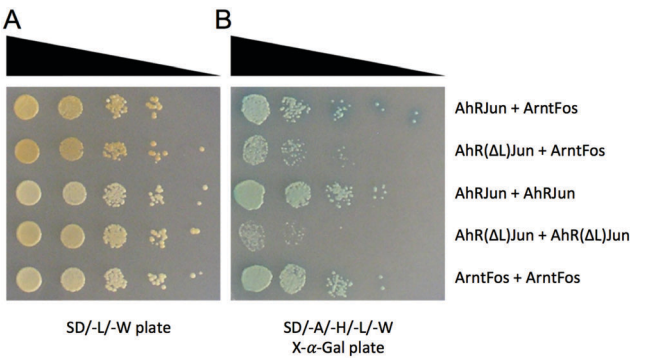


Fig. 3 Y2H assay demonstrating homo- and hetero-dimerization of AhRJun and ArntFos. AhRJun, AhR(ΔL)Jun, and ArntFos were fused to either GAL4DBD or GAL4AD in expression vectors pGBKT7 and pGADT7, respectively. Different combinations of these vectors were transformed into the yeast strain AH109. Transformants were spot-plated in serial dilutions to detect the activation of *HIS3*, *ADE2*, and *MEL1* reporter genes. (A) SD/-L/-W medium, serving as a positive control to show equal plating of cells. (B) SD/-A/-H/-L/-W/X-α-Gal medium. Cells will only grow if their co-expressed DBD- and AD-fusion domains interact.

solution contains 600 nM total monomeric protein—300 nM AhRJun and 300 nM ArntFos. EMSA showed that the dissociation constant (K_d) of the AhRJun/ArntFos solution against the XRE1 site was 337 ± 51 nM (Table 1; representative EMSA titration and binding isotherm shown in Fig. S5, ESI†). The AhRJun/ArntFos solution did not produce any shift in migration of the non-specific control DNA (NS DNA) up to 10 μM protein concentration (Fig. S6a, ESI†), showing XRE1 binding specificity. The Hill coefficient between the AhRJun/ArntFos solution and XRE1 was approximately two, which suggests synergistic and sequential binding of two ligands (protein monomers) to a single target (DNA). Such a protein:DNA binding mechanism coincides with that observed for other native bHLHZ and bZIP proteins such as c-Myc and Max binding to the E-box, and Jun and Fos binding to AP-1:[53–55] XRE1 is bound by one monomer, which cooperatively and rapidly assists the second protein monomer to bind DNA to form the protein dimer:DNA complex that is observed.

In the AhRJun/ArntFos solution, there are three possible dimers that may be responsible for the mobility shift of the XRE1 site: the AhRJun/ArntFos heterodimer, ArntFos homodimer, and AhRJun homodimer. To identify the dimeric species bound to XRE1, we tested AhRJun or ArntFos alone against the site. AhRJun alone did not produce a shift in mobility of XRE1 up to 10 μM protein (Fig. S6c, ESI†). In contrast, ArntFos alone was bound to the XRE1 site with K_d = 1115 ± 307 nM. Although this is a weak interaction between the ArntFos homodimer and XRE1, it is specific, as even 10 μM ArntFos showed no binding to NS DNA (Fig. S6b, ESI†). The interaction between the ArntFos homodimer and XRE1 (K_d = 1115 nM, above), however, was much weaker than that of the AhRJun/ArntFos solution incubated with XRE1 (K_d = 337 nM, above). We therefore eliminated the possibility that the ArntFos homodimer was contributing toward recognition of the XRE1 site, especially given that the AhRJun/ArntFos solution is a mixture in which only half of its total protein is ArntFos, and this mixture incubated with XRE1 gave K_d = 337 nM *vs.* ArntFos alone that gave K_d = 1115 nM. Taken together with the earlier observation that ArntFos alone was unable to produce any signal against XRE1-controlled reporter genes in the MY1H assays, we conclude that the ArntFos homodimer was not involved in the binding reaction between the AhRJun/ArntFos solution and XRE1. The K_d = 337 nM observed represents a specific interaction between the AhRJun/ArntFos heterodimer binding to its intended XRE1 target site.

Table 1 Dissociation constants K_d (nM) of AhRJun & ArntFos with DNA target sites

	XRE1	Arnt E-box	C/EBP	NS
ArntFos/AhRJun heterodimer	337 ± 51	263 ± 19	[a]	[a]
ArntFos homodimer	1115 ± 307	306 ± 11	[b]	[a]
AhRJun homodimer	[a]	[a]	[c]	[a]

Each value is the average of two independent EMSA experiments. Numbers represent the total monomeric protein concentration of each sample.[a] No binding observed at 10 μM protein concentration. [b] Very faint binding observed at 10 μM protein concentration. [c] Smearing of target DNA observed at 5 μM protein concentration.

Apparently, assembly of the AhRJun/ArntFos:XRE1 complex was driven primarily by the DNA target, and not by protein–partner interactions at their dimer interface. In Y2H, where protein/protein interactions are probed in the absence of a DNA target, AhRJun and ArntFos formed comparably stable hetero- and homodimers: in contrast, *when DNA target sites were present*, AhRJun and ArntFos were bound to the XRE1 site exclusively as a heterodimer. This type of DNA-mediated protein heterodimerization has also been observed between the ATF/IGEBP1 heterodimer, driven to heterodimerize by the designed chimeric C/EBP/ATF DNA site, comprising half-sites of C/EBP and ATF.[56] Another example is the folding and assembly of the MyoD/E47 heterodimer that is mediated by the E-box site.[57] Given that AhRJun and ArntFos showed no protein–partner preference, their heterodimerization being driven by the XRE1 site was not surprising. However, these properties of AhRJun and ArntFos contrast with binding of the AP-1 DNA site by the Jun/Fos bZIP heterodimer, where dimerization specificity is primarily driven by protein interactions between Jun and Fos.[58]

The AhRJun/ArntFos heterodimer, rather than ArntFos homodimer, binds to the non-canonical palindromic Arnt E-box site

Given that XRE1 mediated heterodimerization of AhRJun and ArntFos, we then asked how the AhRJun/ArntFos solution would interact with the symmetric Arnt E-box site. We hypothesized that only the ArntFos homodimer would bind to the Arnt E-box, and thus, AhRJun would remain a bystander in this reaction. Previously, the ArntFos homodimer was shown to bind to the Arnt E-box with K_d = 436 ± 107 nM.[35] Assuming the ArntFos homodimer to be the only species binding to Arnt E-box, we expected the K_d for the AhRJun/ArntFos solution against Arnt E-box to be approximately double K_d = 436 nM, *i.e.* K_d = 850–900 nM. Much to our surprise, EMSA showed that the AhRJun/ArntFos solution bound the Arnt E-box with K_d = 263 ± 19 nM, which is not only lower than our prediction, but even lower than the K_d = 337 nM of AhRJun/ArntFos solution against its native XRE1 site.

To explore the discrepancy between the predicted and observed K_d values for the AhRJun/ArntFos solution against the Arnt E-box (850–900 nM *vs.* 263 nM, respectively), we first questioned whether the K_d for ArntFos homodimer:Arnt E-box accurately represented the current system. The K_d of ArntFos:Arnt E-box was originally measured using fluorescence anisotropy, which may produce a different K_d value given the sensitivity of ArntFos towards different experimental conditions.[35] Therefore, we re-determined the K_d using EMSA and obtained K_d = 306 ± 11 nM, which is almost within the range of K_d = 436 ± 107 nM obtained using fluorescence. However, even with the new K_d for ArntFos homodimer:Arnt E-box, we expected the AhRJun/ArntFos solution to bind the Arnt E-box with the ArntFos homodimer at K_d ~ 600 nM, which is still double the K_d between AhRJun/ArntFos solution and Arnt E-box (K_d = 263 nM).

This lead us to examine whether the AhRJun homodimer and/or AhRJun/ArntFos heterodimer was involved in the Arnt E-box mobility shift. A series of qualitative EMSA reactions were performed to investigate these possibilities. First, we found that

10 μM AhRJun alone did not produce any shift in Arnt E-box mobility (Fig. S6c, ESI†), indicating that the AhRJun homodimer does not bind to Arnt E-box. Then, the AhRJun/ArntFos heterodimer was tested using EMSA, by incubating Arnt E-box with either 600 nM ArntFos alone or 600 nM AhRJun/ArntFos solution (Fig. 4, lanes 2 and 3). The ArntFos solution caused a single shift in Arnt E-box mobility, as expected (lane 2), representing the DNA duplex bound by the ArntFos homodimer. In contrast, the reaction with AhRJun/ArntFos solution (lane 3) displayed a single mobility shift that was noticeably slower than that in lane 2. If the ArntFos homodimer was the species responsible for binding to the Arnt E-box, the DNA shifts should have been identical in lanes 2 and 3.

We were therefore left with the only possibility being the AhRJun/ArntFos heterodimer binding to the Arnt E-box in lane 3, Fig. 4. AhRJun and ArntFos have similar molecular weights (~13 kD each) and should have similar bHLHZ structures. Their isoelectric points (pI) differ considerably, however (pI 10 and 8 for AhRJun and ArntFos, respectively). Since our native PAGE running buffer has pH 7.8, the AhRJun/ArntFos:DNA complex would be more positively charged compared to the ArntFos homodimer:DNA complex. This explains the difference in mobilities in lanes 2 and 3: the more positive AhRJun/ArntFos:DNA complex (lane 3) migrated slower through the gel towards the positive electrode than the ArntFos homodimer:DNA complex (lane 2).

To confirm these observations, we repeated the EMSA using ArntFos-SUMO, in which we fused the SUMO tag to ArntFos. Addition of SUMO approximately doubles the molecular weight of ArntFos (SUMO M_W ~ 12 kD), and this mass difference should allow better differentiation between DNA complexes with AhRJun/ArntFos-SUMO (~38 kD) *vs.* the ArntFos-SUMO homodimer (~50 kD). Indeed, the Arnt E-box site incubated with either 600 nM AhRJun/ArntFos-SUMO solution or 600 nM ArntFos-SUMO showed a clear difference in the mobility shift of the Arnt E-box (Fig. 4, lanes 4 and 5). Incubation of the Arnt E-box site with ArntFos-SUMO resulted in a slower migration compared to that of the AhRJun/ArntFos-SUMO solution, as expected. Moreover, only one shift was observed in lanes 3 and 4. For these lanes, any shift caused by ArntFos or ArntFos-SUMO homodimers would have been detected as a second band, running parallel to the bands in lane 2 or 5. Absence of the second band indicates an apparent lack of the ArntFos homodimer interacting with Arnt E-box, which may be due to (1) the ArntFos homodimer not forming at all in the reaction, (2) formation of the ArntFos homodimer, but at a concentration too low to be detected, or (3) formation of the ArntFos homodimer, but its interaction with Arnt E-box being out-competed by the AhRJun/ArntFos heterodimer. Regardless of the mechanism, it is clear that the AhRJun/ArntFos:Arnt E-box complex is the single dominant species present in this reaction.

Taking these results together, we conclude that the AhRJun/ArntFos heterodimer is capable of binding to the non-canonical Arnt E-box site with K_d = 263 nM. Moreover, in the presence of the palindromic Arnt E-box, AhRJun/ArntFos:DNA complex formation is strongly preferred over the ArntFos homodimer binding to DNA, despite the fact that the Arnt E-box site should

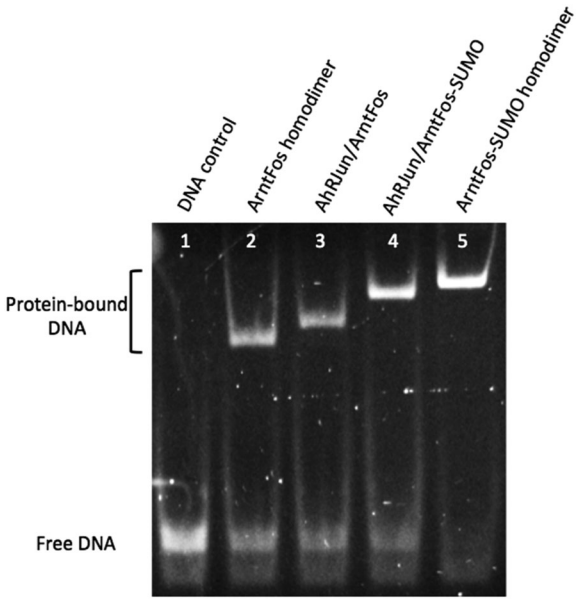


Fig. 4 Only the AhRJun/ArntFos heterodimer forms in the presence of Arnt E-box. The following combinations of proteins were tested for their interactions with the Arnt E-box by EMSA: lane 1, DNA control (no protein); lane 2, ArntFos; lane 3, 1 : 1 AhRJun : ArntFos; lane 4, 1 : 1 AhRJun : ArntFos-SUMO; lane 5, ArntFos-SUMO. Each reaction contained 600 nM total monomeric protein. In all lanes, addition of protein caused a single shift in DNA mobility. The distances of each shift differ in each lane. The difference between lanes 2 and 3 indicates that the mobility shift in lane 3 was caused by the AhRJun/ArntFos heterodimer, and not by the ArntFos homodimer. Similarly, the differences between lanes 4 and 5 indicate that the mobility shift in lane 4 was caused by the AhRJun/ArntFos-SUMO heterodimer, and not by the ArntFos-SUMO homodimer.

be the canonical, symmetric target for the ArntFos homodimer. The K_d value of the AhRJun/ArntFos heterodimer bound to the Arnt E-box was only slightly lower than that of the ArntFos homodimer bound to the Arnt E-box (263 nM *vs.* 306 nM, respectively), which was another reason why we expected to see both hetero- and homodimers bound to Arnt E-box DNA. Interestingly, the opposite was true: the AhRJun/ArntFos:DNA complex was the only species observed, as demonstrated in Fig. 4, lanes 3 and 4.

Max E-box DNA site directs formation of the ArntFos homodimer, not the AhRJun/ArntFos heterodimer

The protein:DNA interaction appears to be the main factor dictating partner selection between AhRJun and ArntFos. If this is the case, the Arnt E-box site directed the formation of the AhRJun/ArntFos heterodimer over the expected "canonical" ArntFos homodimer. To further probe this finding, we repeated EMSA using another E-box variant, Max E-box, which shares the same core E-box sequence CACGTG with the Arnt E-box but is flanked by C/G (Fig. 1). Given the non-canonical Arnt E-box binding by the AhRJun/ArntFos heterodimer, we anticipated that the heterodimer would also bind the Max E-box but with reduced affinity. The Max E-box site was incubated with 600 nM or 1200 nM solutions containing ArntFos alone or AhRJun/ArntFos (Fig. 5, lanes 2–5). The ArntFos homodimer was fairly

weakly bound to the Max E-box (lanes 2 and 3), with $K_d > 600$ nM. This was a weaker interaction compared to that between the ArntFos homodimer and the Arnt E-box (K_d 306 nM), which we expected since the Max E-box is not the preferred target for Arnt. Interestingly, when the Max E-box was titrated with the AhRJun/ArntFos solution, weak binding was detected at 1200 nM (Lane 5), which ran parallel to the shifts caused by the ArntFos homodimer in lanes 2 and 3. Weaker protein:DNA binding observed from the heterodimeric solutions can be ascribed to the fact that only half of the total protein is ArntFos.

These results indicated that the ArntFos homodimer, not the AhRJun/ArntFos heterodimer, caused the shifts in lanes 4 and 5. To confirm that the ArntFos homodimer was bound to Max E-box, EMSA was repeated using ArntFos-SUMO (lanes 6–9). The ArntFos-SUMO homodimer (lanes 8 and 9) bound to the Max E-box site with comparable affinity observed for the ArntFos homodimer, as expected. The reaction between AhRJun/ArntFos-SUMO and the Max E-box site resulted in a single mobility shift parallel to that caused by the ArntFos-SUMO homodimer (Lanes 6–9).

Apparently, the ArntFos *homodimer* is responsible for all DNA binding observed against the Max E-box. This sharply contrasts with the results described between the AhRJun/ArntFos solution and the Arnt E-box site, where the AhRJun/ArntFos *heterodimer* is the sole species binding to the Arnt E-box. Protein hetero- or homodimerization was determined by the DNA target, where the DNA sequence directed the formation of one dimer over others. Although the DNA sequence is expected to influence the formation of protein:DNA complexes, it is intriguing that single base-pair changes *outside* the core DNA target site caused a dramatic shift in the preferences of these interactions.

Two closely related E-box sites direct differential formation of AhRJun and ArntFos dimers

High-resolution structures of AhR bHLH and Arnt bHLH bound to their DNA target would aid our analysis of the interactions between AhRJun, ArntFos, and the E-box variants: in particular, AhR bHLH contacts the two nucleotides that differ between XRE1 and Arnt E-box (XRE1, T**T**G**C**•GTGA; Arnt E-box, T**C**A**C**•GTGA; nucleotide differences are underlined, filled circles separate half-sites). The recently published high-resolution crystal structure of Arnt dimerized with NPAS3 bound to a hypoxia response element variant (HRE, TACGTG) showed the association of the Arnt basic region with the GTG half-site,[59] but no structure for AhR bHLH is available yet. However, possible explanations for the mechanism underlying the protein–protein and protein–DNA interactions observed in this study can still be elucidated. Swanson and colleagues have shown that Arnt binds to the 3′ half-site, GTGA.[20] In the AhRJun/ArntFos heterodimer, the basic region of ArntFos should be capable of forming stable contacts with both XRE1 and Arnt E-box sites, as they share the 3′ GTGA half-site. The basic region of AhR binds to the 5′ TNGC half-site, with preference for T**T**GC or T**C**GC.[19] AhR's preference for cytosine in the second nucleotide explains why it may still bind to the 5′ T**C**AC Arnt E-box half-site.

Furthermore, the **P**SKRH**R** sequence in the AhR basic region is important for half-site recognition.[60] The C-terminally located Arg plays a major role in DNA sequence specificity in other bHLHZ proteins, such as c-Myc, Max, and USF, making these proteins prefer CG nucleotides at the center of the E-box site.[61] The central nucleotides for both XRE1 and the Arnt E-box are CG; AhR could use this Arg to engage similarly with both sites. The N-terminal Pro of the **P**SKRH**R** sequence may come

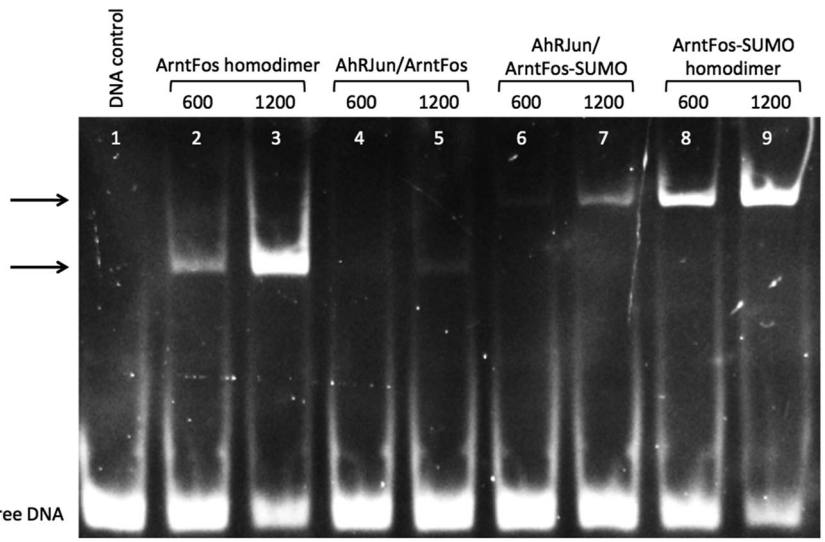


Fig. 5 Only the ArntFos homodimer—not the ArntFos/AhRJun heterodimer—forms in the presence of the Max E-box site. The following combinations of proteins were tested for their interactions with the Max E-box using EMSA: lane 1, DNA control (no protein); lanes 2 and 3, ArntFos homodimer; lanes 4 and 5, 1 : 1 AhRJun : ArntFos; lanes 6 and 7, 1 : 1 AhRJun : ArntFos-SUMO; lanes 8 and 9, ArntFos-SUMO homodimer. Each reaction contained either 600 nM or 1200 nM total monomeric protein. In all lanes, addition of protein caused a single shift in DNA mobility. The migration distances in lanes 2–5 were identical (bottom arrow), indicating that the DNA shifts in lanes 4 and 5 were caused by the ArntFos homodimer, and not by the AhRJun/ArntFos heterodimer. Similarly, the shifts in lanes 6–9 were identical (top arrow), indicating that the shifts in lanes 8 and 9 were caused by the ArntFos-SUMO homodimer, and not the AhRJun/ArntFos-SUMO homodimer.

into contact with the 5′ thymine that is found in both the XRE1 and Arnt E-box sites (Fig. 1);[60] prolines can make favorable interactions with thymines and adenines.[62,63] Taken together, the AhRJun/ArntFos heterodimer appears to have the capacity to bind favorably with most of the Arnt E-box site, except for the adenine at the third nucleotide (TCACGTGA), which is unlikely to be recognized by AhR. These reasons may also explain why the ArntFos homodimer was preferred for binding to the Max E-box that lacks the 5′ thymine, thereby removing one stabilizing interaction that can exist between the AhR basic region and the Arnt E-box. The loss of this interaction may tip the equilibrium in favor of ArntFos homodimer formation at the Max E-box.

The formation of a stable, specific protein dimer:DNA complex is often thought to be the result of an equilibrium reaction, where preformed protein dimers[64] or protein monomers[55,65–68] are constantly associating and dissociating with their cognate and non-cognate DNA sites. This mechanism eventually favors the accumulation of a specific dimer:DNA complex that is the most energetically favorable. In other studies, more than one stable dimer:DNA complex has been shown to form within the same reaction, as in the example of c-Myc/Max and Max/Max hetero- and homo-dimer formation against the E-box site.[20] In contrast, our case illustrates the dynamic nature of dimer:DNA interactions where the target DNA itself can preferentially drive the formation of a single dimeric protein species over others.

The LZ modules in AhRJun and ArntFos primarily serve to stabilize proper folding of the bHLH domains

The dimerization specificity of AhRJun and ArntFos was directed by the target DNA site—not by protein–partner preferences—which was contrary to our original hypothesis. However, the JunD and c-Fos LZs in these hybrid proteins are still essential for protein function. Removing the LZ from AhRJun and ArntFos, thereby leaving the AhRbHLH and ArntbHLH domains by themselves, rendered the proteins inactive in both MY1H and Y2H assays (Fig. S2 and S4, ESI,† rows B, D1 and D2). Previously, we reported similar observations for ArntbHLH, which was capable of strong Arnt E-box binding *in vitro* (K_d = 40 ± 11 nM) but inactive *in vivo* in the Y1H assay.[31] The addition of c-Fos LZ to ArntbHLH (therefore generating ArntFos) restored protein function *in vivo*, although ArntFos had a markedly decreased affinity for the Arnt E-box *in vitro* (K_d = 436 nM,[35] or K_d = 306, this study). Fusing the LZ from bZIP protein C/EBP to ArntbHLH was also enough to restore Arnt E-box binding in *in vivo* yeast assays,[31] and addition of the c-Fos LZ to the bHLH domain of the Max bHLHZ protein restored the helical structure of the otherwise unstructured MaxbHLH domain.[35]

These previous observations, in conjunction with our current results, indicate that these LZs serve as nucleation devices that encourage proper folding of the bHLH domain. Compared to other bHLH and bHLHZ proteins such as E47[69] and Max,[6,7] the ArntbHLH homodimer forms a less hydrophobic four-helix bundle.[10] The inactivity of ArntbHLH, and possibly AhRbHLH, in yeast may stem from the missing PAS domain that helps stabilize the structure. PAS domains have been proposed to initiate folding and/or stabilize dimerization of the bHLH region in a mode similar to that served by the LZ in the bHLHZ Myc/Max heterodimer.[70] However, the fusions of the JunD and c-Fos LZs did not restore the dimerization specificity conferred by the PAS domains in native AhR and Arnt.

In order to further assess the contribution of the JunD and c-Fos LZs, we constructed an AhRJun mutant, AhR(ΔL)Jun, where the first Leu in the Jun LZ was deleted (Fig. 1). As there are 3.6 amino acids per turn in an α-helix, this Leu deletion should offset orientation of the JunD LZ, relative to helix 2 of AhRbHLH, by ~100 degrees, thereby impeding proper dimerization of both the HLH domains and the LZ coiled coil. However, this deletion should not impact the ability of AhRbHLH and ArntbHLH to fold properly. In the Y2H assay, AD- and DBD-AhR(ΔL)Jun were still capable of heterodimerizing with ArntFos to induce reporter gene activation, although AhR(ΔL)Jun alone was incapable of homodimerizing (Fig. S4, ESI† C1–C4 and D3). Furthermore, the Y2H spot assay revealed that the interaction between AD-AhR(ΔL)Jun and DBD-ArntFos was weaker compared to the interaction between AD-AhRJun and DBD-ArntFos (Fig. 3). In addition, a weak interaction between AD-AhR(ΔL)Jun and DBD-AhR(ΔL)Jun was observed in the spot assay, in contrast with the results in Fig. S4 (ESI†). This signal was the weakest of all combinations of proteins tested in the spot assay, which was expected since the two Leu mutations in the AhR(ΔL)Jun homodimer would displace the JunD LZs even farther from each other. It is likely that the serial dilutions of cells used in Fig. S4 (ESI†) (*i.e.* Y2H *HIS3* assay) did not allow for the detection of this weak interaction between AD-AhR(ΔL)Jun and DBD-AhR(ΔL)Jun.

Disruption of the protein/protein interaction by Leu deletion is evidence that JunD and c-Fos LZs contribute towards AhRJun and ArntFos heterodimerization. However, in the MY1H assays, the Leu deletion in AhR(ΔL)Jun did not have any effect on the proteins' ability to interact with DNA. In the *HIS3* assay, cells co-expressing AD-AhR(ΔL)Jun and ArntFos showed growth on SD/-H/-L plates at 30 mM 3-AT (Fig. S3, ESI†). This was confirmed by the ONPG assays, where cells co-expressing AD-AhR(ΔL)Jun and ArntFos showed identical reporter gene activation compared to cells co-expressing AD-AhRJun and ArntFos (Fig. 2, 95.7 ± 3.1 and 95.1 ± 9.0 units of β-galactosidase activity, respectively). Cells expressing only AD-AhR(ΔL)Jun produced a baseline signal. Interestingly, the Leu deletion in AhR(ΔL)Jun impedes its protein/protein interaction (Y2H), but the negative effect vanished once DNA was added as the third interacting factor (MY1H).

We hypothesize that the presence of the XRE1-rich promoter region in the MY1H assay acted as a scaffold on which the AhR(ΔL)Jun/ArntFos heterodimer was able to form. Recognition between the AhR and Arnt basic regions and the DNA major groove provided enough stability to compensate for the influence of the compromised dimerization domain. AhR(ΔL)Jun sheds light on the role of JunD and c-Fos LZs in AhRJun and ArntFos. The LZ domains are essential for our bHLHZ-like proteins to achieve proper structure and function. However, unlike in their native bZIP environment, the primary function of the JunD and c-Fos LZs in AhRJun and ArntFos is to induce proper folding of the bHLH domains, with a much more limited secondary contribution toward protein dimerization.

The AhRJun homodimer does not specifically interact with its "canonical" C/EBP site

The XRE site (TTGC•GTG) is essentially an E-box half-site (GTG) added 3′ to the TTGC half-site of the CCAAT/enhancer-binding protein site (C/EBP).[50] AhRJun and ArntFos are believed to bind to the TTGC and GTG half-sites when bound to XRE1 as a heterodimer. ArntFos alone, much like the native Arnt bHLH/PAS domain, binds to the E-box site (CACGTG) as a homodimer.[21,35] We therefore asked if AhRJun alone, which binds to the TTGC half site, was capable of binding to the C/EBP site TTGCGCAA as a homodimer. In EMSA, AhRJun alone did not produce a clear shift in mobility at the C/EBP, XRE1, Arnt E-box, or NS DNA sites, even at 10 μM protein (Fig. S6c, ESI†). At concentrations exceeding 5 μM, AhRJun appeared to cause DNA to precipitate in the wells, indicating unstable, nonspecific protein:DNA aggregation. In nature, both AhR and JunD strongly prefer heterodimerization over homodimerization,[9,46] and to our knowledge, there has been no report of AhR or its variants being capable of DNA binding as a homodimer. Although the Y2H assay showed that AhRJun is capable of homodimerizing, the AhRJun homodimer did not bind to the expected C/EBP DNA target.

Conclusion

Our work demonstrates the dynamic nature inherent for protein:DNA interactions and sheds light on the magnitude of the influence that the target DNA sequence exerts during the formation of protein:DNA species. We successfully generated a bHLHZ-like hybrid protein *heterodimer* capable of binding specifically to target DNA. However, protein dimerization during DNA binding was primarily driven by the target DNA site, not by protein/protein interactions. Therefore, in some systems, the DNA sequence not only influences the affinity of protein:DNA interaction, but also can directly impact the protein partner selection by preferring a single protein dimer over other possibilities. We emphasize that in our example, the E-box target site essentially controlled protein–partner selection through subtle differences in the nucleotide core sequence: modification of the E-box sequence itself was not necessary to cause a complete shift in preference between protein hetero- and homo-dimer formation.

Our findings also have implications on the evolutionary origin of the bHLH superfamily of proteins. Based on a study of bHLH domains in 242 bHLH superfamily members, Atchley and Fitch produced a phylogenetic tree to classify family members according to evolutionary relationships, and found that their analyses could not rigorously support the hypothesis of a single evolutionary origin for the bHLH domain.[71] However, a further intensive investigation of the non-bHLH components of 122 bHLH superfamily members provided strong evidence that supports the hypothesis that bHLH proteins have undergone modular evolution by domain shuffling, a process that involves domain insertion and rearrangement.[72] Our domain swapping experiment provides additional, complementary evidence for the hypothesis of modular evolution by domain shuffling.

Consistent with previous experiments,[9,29,30,32,35] our studies revealed that individual domains of bHLH and bZIP proteins are swappable. In addition, the influence of the DNA sequence driving the preferential formation of particular protein dimers indicates the possible role of DNA during diversification of DNA-binding proteins: the PAS domains in the bHLH/PAS AhR/Arnt heterodimer could be replaced by the LZ domains of bZIP, a different transcription factor family from bHLH, while maintaining dimerized protein structure and specific DNA-binding function. These LZ domains in AhRJun and ArntFos, however, contributed to DNA binding through a different functionality compared to their native mechanism in the bZIP, indicating the capacity of these domains to diversify their roles during domain shuffling. Taken together, our results provide further evidence that domain shuffling is likely the evolutionary pathway that provided diverse structural patterns and variable functions in the bHLH superfamily, and co-evolution of the bHLH superfamily with the DNA target site may have heavily impacted the outcome of diversification of these proteins.

Competing financial interest

The authors declare no competing financial interest.

Acknowledgements

The authors are grateful for funding from an NSERC Discovery Grant and the Collaborative Health Research Program (CHRP).

References

1 T. D. Littlewood and G. I. Evan, *Protein Profile*, 1994, **1**, 639–701.
2 S. Jones, *Genome Biol.*, 2004, **5**, 226.
3 R. J. Kewley, M. L. Whitelaw and A. Chapman-Smith, *Int. J. Biochem. Cell Biol.*, 2004, **36**, 189–204.
4 C. Murre, G. Bain, M. A. Vandijk, I. Engel, B. A. Furnari, M. E. Massari, J. R. Matthews, M. W. Quong, R. R. Rivera and M. H. Stuiver, *Biochim. Biophys. Acta, Gene Struct. Expression*, 1994, **1218**, 129–135.
5 M. E. Massari and C. Murre, *Mol. Cell. Biol.*, 2000, **20**, 429–440.
6 A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff and S. K. Burley, *Nature*, 1993, **363**, 38–45.
7 P. Brownlie, T. A. Ceska, M. Lamers, C. Romier, G. Stier, H. Teo and D. Suck, *Structure*, 1997, **5**, 509–520.
8 S. K. Nair and S. K. Burley, *Cell*, 2003, **112**, 193–205.
9 A. Chapman-Smith and M. L. Whitelaw, *J. Biol. Chem.*, 2006, **281**, 12535–12545.
10 J. L. Huffman, A. Mokashi, H. P. Bachinger and R. G. Brennan, *J. Biol. Chem.*, 2001, **276**, 40537–40544.
11 S. Cuthill, A. Wilhelmsson and L. Poellinger, *Mol. Cell. Biol.*, 1991, **11**, 401–411.
12 M. Whitelaw, I. Pongratz, A. Wilhelmsson, J.-A. Gustafsson and L. Poellinger, *Mol. Cell. Biol.*, 1993, **13**, 2504–2514.
13 L. Wu and J. P. Whitlock, *Nucleic Acids Res.*, 1993, **21**, 119–125.

14 A. Fujisawa-Sehara, K. Sogawa, M. Yamane and Y. Fujii-Kuriyama, *Nucleic Acids Res.*, 1987, **15**, 4179–4191.
15 T. Schmid, J. Zhou and B. Brüne, *J. Cell. Mol. Med.*, 2004, **8**, 423–431.
16 S. G. Bacsi, S. Reisz-Porszasz and O. Hankinson, *Mol. Pharmacol.*, 1995, **47**, 432–438.
17 J. C. Rowlands and J.-Å. Gustafsson, *Crit. Rev. Toxicol.*, 1997, **27**, 109–134.
18 H. I. Swanson, W. K. Chan and C. A. Bradfield, *J. Biol. Chem.*, 1995, **270**, 26292–26302.
19 S. Li, X. Pei, W. Zhang, H. Q. Xie and B. Zhao, *Int. J. Mol. Sci.*, 2014, **15**, 6475–6487.
20 H. I. Swanson and J.-H. Yang, *Nucleic Acids Res.*, 1999, **27**, 3205–3212.
21 F. Wang, S. Shi, R. Zhang and O. Hankinson, *Biol. Chem.*, 2006, **387**, 1215–1218.
22 I. Pongratz, C. Antonsson, M. L. Whitelaw and L. Poellinger, *Mol. Cell. Biol.*, 1998, **18**, 4079–4088.
23 T. Ellenberger, *Curr. Opin. Struct. Biol.*, 1994, **4**, 12–21.
24 W. H. Landschulz, P. F. Johnson and S. L. McKnight, *Science*, 1988, **240**, 1759–1764.
25 P. B. Harbury, T. Zhang, P. S. Kim and T. Alber, *Science*, 1993, **262**, 1401–1407.
26 A. Acharya, V. Rishi, J. Moll and C. Vinson, *J. Struct. Biol.*, 2006, **155**, 130–139.
27 L. J. Ransone and I. M. Verma, *Annu. Rev. Cell Biol.*, 1990, **6**, 539–557.
28 P. Angel and M. Karin, *Biochim. Biophys. Acta, Gen. Subj.*, 1991, **1072**, 129–157.
29 L. J. Ransone, P. Wamsley, K. L. Morley and I. M. Verma, *Mol. Cell. Biol.*, 1990, **10**, 4565–4573.
30 K. Struhl, *Cell*, 1987, **50**, 841–846.
31 H.-K. Chow, J. Xu, S. H. Shahravan, A. T. De Jong, G. Chen and J. A. Shin, *PLoS One*, 2008, **3**, e3514.
32 J. Xu, G. Chen, A. T. De Jong, S. H. Shahravan and J. A. Shin, *J. Am. Chem. Soc.*, 2009, **131**, 7839–7848.
33 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.
34 F. Ahmadpour, R. Ghirlando, A. T. De Jong, M. Gloyd, J. A. Shin and A. Guarne, *PLoS One*, 2012, **7**, e32136.
35 G. Chen, A. T. De Jong and J. A. Shin, *Mol. BioSyst.*, 2012, **8**, 1286–1296.
36 G. Chen and J. A. Shin, *Anal. Biochem.*, 2008, **382**, 101–106.
37 G. Chen, L. M. DenBoer and J. A. Shin, *BioTechniques*, 2008, **45**, 295–304.
38 C. T. Chung, S. L. Niemela and R. H. Miller, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 2172–2175.
39 R. J. Dohmen, A. W. M. Strasser, C. B. Höner and C. P. Hollenberg, *Yeast*, 1991, **7**, 691–692.
40 J. Tornow and G. M. Santangelo, *Gene*, 1990, **90**, 79–85.
41 B. Titz, S. Thomas, S. V. Rajagopala, T. Chiba, T. Ito and P. Uetz, *Nucleic Acids Res.*, 2006, **34**, 955–967.
42 E. Izumchenko, M. Wolfson, E. A. Golemis and I. G. Serebriiskii, in *Yeast Gene Analysis*, ed. I. Stansfield and M. J. R. Stark, Elsevier Academic Press Inc., San Diego, 2nd edn, 2007.
43 G. H. Bird, A. R. Lajmi and J. A. Shin, *Biopolymers*, 2002, **65**, 10–20.
44 Y. Xie and D. B. Wetlaufer, *Protein Sci.*, 1996, **5**, 517–523.
45 I.-S. Chan, A. V. Fedorova and J. A. Shin, *Biochemistry*, 2007, **46**, 1663–1671.
46 J. R. S. Newman and A. E. Keating, *Science*, 2003, **300**, 2097–21010.
47 S. G. Bacsi and O. Hankinson, *J. Biol. Chem.*, 1996, **271**, 8843–8850.
48 M. R. Probst, S. Reisz-Porszasz, R. V. Abunag, M. S. Ong and O. Hankinson, *Mol. Pharmacol.*, 1993, **44**, 511–518.
49 F. Saatcioglu, D. J. Perry, D. S. Pasco and J. B. Fagan, *Mol. Cell. Biol.*, 1990, **10**, 6408–6416.
50 C. R. Vinson, P. B. Sigler and S. L. McKnight, *Science*, 1989, **246**, 911–916.
51 J. Estojak, R. Brent and E. A. Golemis, *Mol. Cell. Biol.*, 1995, **15**, 5820–5829.
52 H. I. Swanson, *Chem. – Biol. Interact.*, 2002, **141**, 63–76.
53 A. Banerjee, H. Hu and D. J. Goss, *Biochemistry*, 2006, **45**, 2333–2338.
54 J. Hu, A. Banerjee and D. J. Goss, *Biochemistry*, 2005, **44**, 11855–11863.
55 J. J. Kohler and A. Schepartz, *Biochemistry*, 2001, **40**, 130–142.
56 C. R. Vinson, T. W. Hai and S. M. Boyd, *Genes Dev.*, 1993, **7**, 1047–1058.
57 H. Wendt, R. M. Thomas and T. Ellenberger, *J. Biol. Chem.*, 1998, **273**, 5735–5743.
58 P. Lamb and S. L. McKnight, *Trends Biochem. Sci.*, 1991, **16**, 417–422.
59 D. Wu, X. Su, N. Potluri, Y. Kim and F. Rastinejad, *eLife*, 2016, **5**, e18790.
60 H. I. Swanson and J.-H. Yang, *J. Biol. Chem.*, 1996, **271**, 31657–31665.
61 C. V. Dang, C. Dolde, M. L. Gillison and G. J. Kato, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 599–602.
62 N. C. Seeman, J. M. Rosenberg and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 804–808.
63 R. Rohs, S. M. West, A. Sosinsky, P. Q. Liu, R. S. Mann and B. Honig, *Nature*, 2009, **461**, 1248–1253.
64 S. Sauve, J.-F. Naud and P. Lavigne, *J. Mol. Biol.*, 2007, **365**, 1163–1175.
65 J. J. Kohler, S. J. Metallo, T. L. Schneider and A. Schepartz, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 11735–11739.
66 S. Cranz, C. Berger, A. Baici, I. Jelesarov and H. R. Bosshard, *Biochemistry*, 2004, **43**, 718–727.
67 C. Berger, I. Jelesarov and H. R. Bosshard, *Biochemistry*, 1996, **35**, 14984–14991.
68 C. Berger, L. Piubelli, U. Haditsch and H. R. Bosshard, *FEBS Lett.*, 1998, **425**, 14–18.
69 T. Ellenberger, D. Fass, M. Arnaud and S. C. Harrison, *Genes Dev.*, 1994, **8**, 970–980.
70 A. Chapman-Smith, J. K. Lutwyche and M. L. Whitelaw, *J. Biol. Chem.*, 2004, **279**, 5353–5362.
71 W. R. Atchley and W. M. Fitch, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 5172–5176.

72 B. Morgenstern and W. R. Atchley, *Mol. Biol. Evol.*, 1999, **16**, 1654–1663.
73 D. Wu, X. Su, N. Potluri, Y. Kim and F. Rastinejad, *eLife*, 2016, **5**, e18790.
74 J. N. M. Glover and S. C. Harrison, *Nature*, 1995, **373**, 257–261.
75 D. Wang, H. Hashimoto, X. Zhang, B. G. Barwick, S. Lonial, L. H. Boise, P. M. Vertino and X. Cheng, *Nucleic Acids Res.*, 2016, DOI: 10.1093/nar/gkw1184.
76 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.