

The bZIP Targets Overlapping DNA Subsites within a Half-Site, Resulting in Increased Binding Affinities[†]

I-San Chan,[‡] S. Hesam Shahravan,[‡] Anna V. Fedorova,[‡] and Jumi A. Shin^{‡,§,*}

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

Received February 29, 2008; Revised Manuscript Received April 29, 2008

ABSTRACT: We previously reported that the wt bZIP, a hybrid of the GCN4 basic region and C/EBP leucine zipper, not only recognizes GCN4 cognate site AP-1 (TGA³CTCA) but also selectively targets noncognate DNA sites, in particular the C/EBP site (TTGCGCAA). In this work, we used electrophoretic mobility shift assay and DNase I footprinting to investigate the factors driving the high affinity between the wt bZIP and the C/EBP site. We found that on each strand of the C/EBP site, the wt bZIP recognizes two 4 bp subsites, TTGC and TGCG, which overlap to form the effective 5 bp half-site (TTGCG). The affinity of the wt bZIP for the overall 5 bp half-site is ≥ 10 -fold stronger than that for either 4 bp subsite. Our results suggest that interactions of the wt bZIP with both subsites contribute to the strong affinity at the overall 5 bp half-site and, consequently, the C/EBP site. Accordingly, we propose that the wt bZIP undergoes conformational changes to slide between the two overlapping subsites on the same DNA strand and establish sequence-selective contacts with the different subsites. The proposed binding mechanism expands our understanding of what constitutes an actual DNA target site in protein–DNA interactions.

The basic region/leucine zipper (bZIP)¹ is the simplest DNA-binding motif used by transcription factors. Complexes of the GCN4 bZIP with the cognate AP-1 and CRE sites show how this motif engages sequence-specific DNA binding (1–5). The bZIP targets DNA as a dimer of short, seamless α -helices: each monomer comprises a basic region for targeting the DNA major groove and a leucine zipper for dimerization via coiled-coil structure. Thus, the bZIP provides a straightforward, native motif for examination of the relationship between protein structure and DNA-binding function.

We previously generated the wt bZIP (wild type), a hybrid of the GCN4 basic region and C/EBP leucine zipper that maintains α -helical structure and DNA-binding function comparable to those of the native GCN4 bZIP (Figure 1) (6, 7). The GCN4 basic region is responsible for targeting

the 4 bp cognate half-site TGAC, and the C/EBP leucine zipper for protein dimerization. Similar to native transcription factors, the wt bZIP can tolerate sequence variation in DNA target sites. For example, both the wt bZIP and the native GCN4 bZIP bind with similar affinities to the AP-1 and CRE sites, which differ by one centrally located base pair (7, 8). The wt bZIP also selectively targets some noncognate sites, including the C/EBP site (TTGCGCAA, eponymous cognate site of CCAAT/enhancer binding protein) and Arnt E-box site (TCACGTGA, core enhancer box underlined with immediate flanking bases preferred by bHLH/PAS protein Arnt) (9, 10).

Interestingly, at the C/EBP site, the wt bZIP demonstrates a significant affinity only 100-fold reduced from that for binding at the AP-1 and CRE sites; in contrast, affinities at other noncognate sites, such as the Arnt E-box site, are ≥ 2000 -fold reduced (9). Additionally, the half-site binding affinities at the TTGC and TCAC sequences are the same; however, the wt bZIP targets the C/EBP site (abutting TTGC sequences) with 20-fold stronger affinity than the Arnt E-box site (abutting TCAC sequences) (Figure 1) (10). These observations prompted us to explore further the determinants of affinity for the binding between the wt bZIP and DNA targets.

We therefore designed a number of target sites based on the C/EBP site and used DNase I footprinting and electrophoretic mobility shift assay (EMSA) to examine their interactions with the wt bZIP. We found that on each strand of the C/EBP site, the wt bZIP selectively targets not one but two overlapping 4 bp subsites, TTGC and TGCG, accessible by a 1 bp shift, resulting in an overall 5 bp TTGCG half-site. By accessing both subsites, the wt bZIP exhibits increased binding affinity at the overall half-site.

[†] We express gratitude for funding from the National Institutes of Health (RO1GM069041), the Canadian Foundation for Innovation/Ontario Innovation Trust (CFI/OIT), the Premier's Research Excellence Award (PREA), and the University of Toronto.

* To whom correspondence should be addressed. Phone: (905) 828-5355. Fax: (905) 828-5425. E-mail: jumi.shin@utoronto.ca.

[‡] Department of Chemistry, University of Toronto.

[§] Institute of Biomaterials and Biomedical Engineering, University of Toronto.

¹ Abbreviations: bZIP, basic region/leucine zipper; CRE, cAMP-response element; C/EBP, CCAAT/enhancer binding protein; Arnt, aryl hydrocarbon receptor nuclear translocator; E-box, enhancer box; bHLH/PAS, basic/helix–loop–helix/Per-Arnt-Sim; EMSA, electrophoretic mobility shift assay; ESI-MS, electrospray ionization mass spectrometry; e-wt bZIP, bacterially expressed wt bZIP; s-wt bZIP, chemically synthesized wt bZIP; HPLC, high-performance liquid chromatography; BSA, bovine serum albumin; DTT, dithiothreitol; T-leap, temperature leap; PAGE, polyacrylamide gel electrophoresis; EDTA, ethylenediaminetetraacetic acid; TBE, Tris-borate-EDTA; K_d , apparent dimeric equilibrium dissociation constant; θ_{app} , bound DNA fraction; $\Delta\theta_{app}$, net bound DNA fraction; NMR, nuclear magnetic resonance.

A. Sequences of the wt bZIP

Bacterially Expressed wt bZIP (e-wt bZIP, 96-mer)
MGGSHHHHHHGMASMTGGQQMGRDLYDDDDK-
DPAALKRARNTEAARRSRARKLQRMKQ-LEQKVLELT
 SDNDRLRKRVEQLSRELDLT-**GGCGGYYYYY**
 Chemically Synthesized wt bZIP (s-wt bZIP, 57-mer)
DPAALKRARNTEAARRSRARKLQRMKQ-LEQKVLELT
 SDNDRLRKRVEQLSRELDLT-**Y**

B. Sequences used in footprinting analysis

AP-1 5'-TCCCTCCGGA**TGACTCA**TTTTTGGCA-3'
 C/EBP 5'-CCTGCAGGAA**TTGCGCAA**TGAAGGTT-3'
 C/EBP-2 5'-CCTGCAGGAA**TTGCGCA**TGAAGGTT-3'
 5H-LR 5'-CCTGCAGGAA**TTGCGT**TTGAAGGTT-3'

C. Sequences used in EMSA

AP-1 5'-TGCAGGAA**TGACTCA**TTGAAGGTT-3'
 C/EBP 5'-TGCAGGAA**TTGCGCAA**TGAAGGTT-3'
 C/EBP-2 5'-TGCAGGAA**TTGCGCA**TGAAGGTT-3'
 R-F 5'-TGCAGGAA**TGCGCGCA**TGAAGGTT-3'
 AP-1H24 5'-TGCAGGAA**TGACT**TTTTGAAGGTT-3'
 5H-LR 5'-TGCAGGAA**TTGCGT**TTGAAGGTT-3'
 R-H 5'-TGCAGGAA**TTGCGT**TTGAAGGTT-3'
 Arnt E-box 5'-TGCAGGAA**TCACG**TGAAGGTT-3'
 NS 5'-TGCAGGAA**AGGAA**TTCC**TGAAGGTT**-3'
 NS2 5'-TGCAGGAA**AAAA**TGCTGAAGGTT-3'
 Arnt E-box half 5'-TGCAGGAA**TCAC**TGAAGGTT-3'
 AP-1 half 5'-TGCAGGAA**TGACT**TGAAGGTT-3'
 L-H 5'-TGCAGGAA**TTGCG**TGAAGGTT-3'

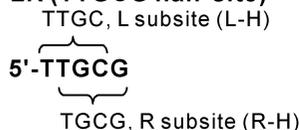
D. 5H-LR (TTGCG half-site)

FIGURE 1: (A) Sequences of the wt bZIP. The bacterially expressed wt bZIP includes an N-terminal sequence with a six-His tag from expression vector pTrcHis B (31 amino acids, bold, underlined), the GCN4 basic region (27 amino acids, italicized), the C/EBP leucine zipper (29 amino acids), and a C-terminal linker (9 amino acids, bold, underlined) (6). The N-terminal Met was cleaved during post-translational modification (11). The chemically synthesized wt bZIP includes the same bZIP domain with a C-terminal Tyr (bold, underlined). (B) Sequences used in DNase I footprinting analysis. Only core target sequences and surrounding flanking sequences are shown. Core target sequences are in bold, and the entire inserted sequences between flanking sequences are underlined. Flanking sequences of C/EBP, C/EBP-2, and 5H-LR are identical to those for EMSA. Flanking sequences of the AP-1 duplex are from bp -87 to -102 of the *his3* promoter region of the yeast genome (22, 23). (C) Sequences Used in EMSA analysis (24 or 20 bp duplexes). Core target sequences are in bold, and the entire sequences inserted between flanking sequences are underlined. Identical flanking sequences were chosen to minimize DNA secondary structure, and the core target sequences of 5H-LR and AP-1H24 were shifted for the same reason. The two thymines at the 3'-end of each duplex were ³²P-labeled. (D) Construction of the 5H-LR site. The 5H-LR site contains two 4 bp sequences. The TTGC sequence is at the 5'-end (L subsite); the TGCG sequence is at the 3'-end (R subsite).

We therefore propose a binding mechanism that offers further insight into what constitutes the DNA target site in protein-DNA interactions.

EXPERIMENTAL PROCEDURES

The following protocols have been described in detail previously (6, 9, 10); the following is a brief summary.

Preparation of wt bZIP Proteins. The expressed wt bZIP [e-wt bZIP (Figure 1)] was cloned into pTrcHis B (Invitrogen) and expressed in *Escherichia coli* strain BL21(DE3)

(Stratagene). The six-His-tagged e-wt bZIP was initially purified with TALON cobalt metal ion affinity resin (Clontech) (6). The synthetic wt bZIP (s-wt bZIP) was produced by standard Fmoc chemistry (Biomer Technologies). Both versions of the wt bZIP were purified by HPLC (Beckman System Gold) on a reversed-phase C₁₈ column (Vydac). The wt bZIP identity was verified by ESI-MS (Waters Micromass ZQ model MM1) (9); the N-terminal Met of the e-wt bZIP was cleaved during post-translational modification (11).

DNase I Footprinting Analysis. The core DNA targets and their surrounding flanking sequences (Figure 1) were cloned into pUC19 between restriction sites *Bam*H I and *Sac* I. DNA duplexes (~650 bp) were radiolabeled with ³²P at the 5'- or 3'-termini (8, 10). The lyophilized e-wt bZIP was freshly dissolved to required concentrations in TKMC buffer [20 mM Tris (pH 7.5), 4 mM KCl, 2 mM MgCl₂, 1 mM CaCl₂, 100 μg/mL nonacetylated BSA, 5 mM DTT, 1 μg/mL poly(dI-dC), and 5% glycerol]. To prevent e-wt bZIP aggregation, the temperature-leap (*T*-leap) tactic was used (12): protein solutions were incubated at 4 °C for ≥2 h and at 37 °C for 1 h, followed by addition of ³²P-labeled DNA fragment (~5000 cpm/reaction, DNA concentrations of <15 pM), and further incubation at 4 °C for 2 h, 37 °C for 1 h, and 22 °C for 30 min. DNase I digestion occurred for 3 min at 22 °C. Digestion with the e-wt bZIP at monomeric concentrations of 0, 0.5, and 3 μM required DNase I at 0.01, 0.05, and 0.2 mg/mL, respectively. DNase I cleavage activity decreased at increasing wt bZIP concentrations, possibly due to soluble aggregates of the wt bZIP and DNase I (9). Reaction products were separated by denaturing 8% PAGE; dried gels were analyzed by phosphorimaging (Molecular Dynamics Storm 840 and ImageQuant, version 5.2). All experiments were repeated at least three times.

Electrophoretic Mobility Shift Assay (EMSA). DNA duplexes were radiolabeled with ³²P at the 3'-terminus (Figure 1). The lyophilized wt bZIP was freshly dissolved to a monomeric concentration of 2 μM with EMSA buffer [20 mM Tris, 1 mM phosphate (pH 7.5), 5 mM NaCl, 4 mM KCl, 2 mM MgCl₂, 1 mM CaCl₂, 1 mM EDTA, 100 μg/mL nonacetylated BSA, 2 μg/mL poly(dI-dC), 200 mM guanidine-HCl, and 10% glycerol] and sequentially diluted to the required concentrations. Protein solutions were heated at 90 °C for 10 min and cooled to 22 °C over 4 h, followed by addition of ³²P-labeled DNA duplex (~3000 cpm/reaction, DNA final concentrations of <15 pM), and further incubated at 4 °C overnight, 37 °C for 1 h, and 22 °C for 30 min. Reaction mixtures were analyzed by native 10% PAGE at 120 V and 22 °C for 3 h in qualitative experiments and for 90 min in thermodynamic titrations. Gels were analyzed as described above. All experiments were repeated at least three times.

Although the *T*-leap was not typically necessary for the s-wt bZIP at <1.0 μM, we consistently applied the *T*-leap to all experiments. Several different *T*-leaps were tried, with the best results from those described above. EMSA was also tried with TKMC buffer and a higher-NaCl buffer [15 mM Tris (pH 7.5), 75 mM NaCl, 1.35 mM KCl, 5 mM MgCl₂, and 2.5 mM CaCl₂]. The wt bZIP has poorer solubility in TKMC and exhibits band broadening in the higher-NaCl buffer (9, 10). Also, we varied the concentration of poly(dI-dC) between 2 and 60 μg/mL and found no effect on binding activity (9).

Table 1: Dissociation Constants and Net Bound DNA Fractions for the Synthetic wt bZIP in Complex with Target Sites^f

Target Site	Sequence	Half-Sites/Subsites		K_d (M ²) ^a	$\Delta\theta_{app}$ ^b	
		Coding Strand	Complementary Strand		200 nM	1000 nM
Arnt E-box	TCACGTGA	TCAC	TCAC	$3.2 \pm 0.0 \times 10^{-13d,e}$	0.10 ± 0.00^e	0.36 ± 0.03^e
Arnt E-box Half	TCAC	TCAC		$>1.0 \times 10^{-11e}$	NA ^{c,e}	0.02 ± 0.00^e
C/EBP	TTGCGCAA	TTGC, TGCG	TTGC, TGCG	$1.5 \pm 0.4 \times 10^{-14d,e}$	0.35 ± 0.07^e	0.61 ± 0.05^e
L-H	TTGC	TTGC		$>1.0 \times 10^{-11e}$	NA ^{c,e}	0.02 ± 0.00^e
5H-LR	TTGCG	TTGC, TGCG		$1.1 \pm 0.4 \times 10^{-12}$	0.03 ± 0.00	0.20 ± 0.02
R-H	TGCG	TGCG		$>>1.0 \times 10^{-11}$	NA ^c	NA ^c
C/EBP-2	TGCGCA	TGCG	TGCG	$2.0 \pm 0.1 \times 10^{-12}$	0.02 ± 0.00	0.15 ± 0.01
R-F	TGCGCGCA	TGCG	TGCG	$>1.0 \times 10^{-11}$	0.01 ± 0.00	0.03 ± 0.02

^a Each reported dissociation constant (K_d) is the average of two values from independent data sets of independent thermodynamic titration experiments. Only data sets fit to eq 1 with R values of >0.970 are used for calculation of the reported K_d values. ^b Net bound DNA fractions ($\Delta\theta_{app}$) are given as references of binding affinities (see Experimental Procedures). Each $\Delta\theta_{app}$ is the average of two values from the same independent data sets for K_d determination. The $\Delta\theta_{app}$ values were measured with the wt bZIP at a monomeric concentration of 200 or 1000 nM. ^c No dimeric wt bZIP DNA-binding activity observed in titration experiments. ^d Monomeric dissociation constants or $\Delta\theta_{app}$ values previously published in ref 9. ^e Monomeric dissociation constants or $\Delta\theta_{app}$ values previously published in ref 10. ^f Thermodynamic EMSA titrations were performed with the s-wt bZIP, the solubility of which is higher than that of the e-wt bZIP.

Determination of Dimeric K_d Values. Apparent dimeric dissociation constants (K_d) were measured by thermodynamic EMSA titrations and were determined by fitting the bound DNA fraction (θ_{app}) values versus monomeric wt bZIP concentrations ($[M]$) to eq 1 (9, 10, 13, 14). θ_{app} is the overall intensity of the bound DNA band divided by the sum of the overall intensity of the free and bound DNA bands.

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

where θ_{min} is the bound DNA fraction in the absence of the wt bZIP and θ_{max} is the bound DNA fraction when DNA binding is saturated. Only data sets fit to eq 1 with R values of >0.970 are reported. Each K_d is the average of two values from independent data sets (Table 1).

Net bound DNA fractions ($\Delta\theta_{app}$) at specified monomeric concentrations of the wt bZIP are defined as the net increases of θ_{app} from θ_{min} (10). Thus

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

According to eq 2, at specific wt bZIP concentrations, the lower the K_d , the higher the $\Delta\theta_{app}$. Therefore, $\Delta\theta_{app}$ values serve as references for binding affinities.

RESULTS

We initially observed that the wt bZIP displays significant affinity for the C/EBP site only 100-fold weaker than that at the cognate AP-1 and CRE sites [both sites give a K_d of 1.6×10^{-16} M²; previous data recalculated as dimeric dissociation constants (10)]. In contrast, the affinities at other noncognate sites, including the Arnt E-box site, are reduced ≥ 2000 -fold (Table 1) (9). Furthermore, the affinity of the wt bZIP at the Arnt E-box site increases ~ 30 -fold from that at its half-site; likewise, the affinity of the wt bZIP at the AP-1 site increases 44-fold from that at its half-site ($K_d = 7.0 \times 10^{-15}$ M²) (10). These affinity increases are in agreement with measurements by Hollenbeck and Oakley; the affinity of their GCN4 bZIP derivative at the AP-1 and CRE sites increases 30- and 80-fold, respectively, from that at the TGAC half-site (14). In contrast, compared with the half-site binding affinity of the wt bZIP at the TTGC sequence, the full-site binding affinity increases ≥ 700 -fold at the C/EBP site (abutting TTGC sequences) (10). These

observations prompted our investigation of the factors driving the strong affinity of the wt bZIP at the C/EBP site.

Can the wt bZIP target other sequences, aside from the TTGC sequence, within the C/EBP site? At the C/EBP site, the TTGC sequence [termed ‘‘L’’ for left (Figure 1)] is located at the 5'-end, and the TGCG sequence (‘‘R’’ for right) is shifted by 1 bp toward the 3'-end. We considered whether the R sequence contributes to the strong affinity of the wt bZIP at the C/EBP site. Thus, we designed four DNA target sites: (1) the C/EBP-2 site containing complementary R sequences identically situated as in the C/EBP site, (2) the R-F site (‘‘right-full’’) comprising abutting complementary R sequences, (3) the R-H site (‘‘right-half’’) containing only one R sequence, and (4) the 5H-LR site comprising the overall 5 bp TTGCG half-site with overlapping L and R sequences.

We analyzed wt bZIP–DNA interactions with EMSA and DNase I footprinting with expressed and synthetic versions of the wt bZIP [e- and s-wt bZIP (Figure 1)], both of which display comparable DNA-binding function, as shown previously (10). Qualitative EMSA performed with the e- and s-wt bZIP produced the same results. Therefore, DNase I footprinting, performed with the e-wt bZIP for direct comparison with previously published data with the same version of the wt bZIP, is directly comparable with EMSA data gained with the e- or s-wt bZIP (9, 10). To determine apparent dimeric dissociation constants (K_d) and net bound DNA fractions ($\Delta\theta_{app}$), thermodynamic EMSA titrations were performed with the s-wt bZIP, as it is less prone to aggregation than the e-wt bZIP (9, 10). To allow direct comparison of binding affinities, target sites used in EMSA are flanked by the same sequences.

DNase I Footprinting and EMSA. Clear footprints are observed between the e-wt bZIP and the C/EBP-2 and 5H-LR sites (Figure 2; see Figure S2 for quantitative phosphorimaging analysis); footprints at the AP-1 and C/EBP sites were presented previously and are repeated in this study as controls (9, 10). The AP-1, C/EBP, C/EBP-2, and 5H-LR sites are situated identically within the DNA duplexes (Figure 1), and the footprints and surrounding hypersensitivity cleavage regions are located like those for the AP-1 and C/EBP controls. Thus, DNase I footprinting shows that the wt bZIP selectively targets the C/EBP-2 and 5H-LR sites.

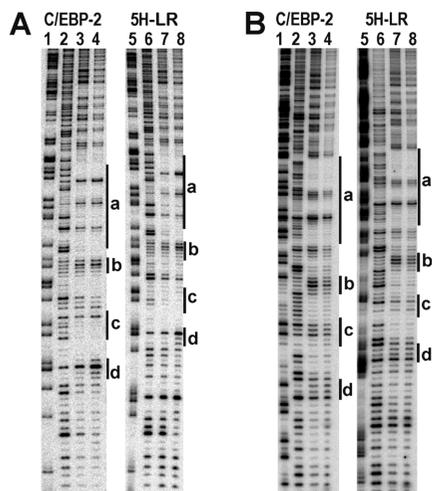


FIGURE 2: DNase I footprinting analysis of the e-wt bZIP targeting the C/EBP-2 and 5H-LR sites: (A) 5' ^{32}P -end-labeled DNA and (B) 3' ^{32}P -end-labeled DNA. Data presented in panel A are from a single gel with separations and labeling provided for clarity; likewise for panel B. Lanes 1 and 5: chemical sequencing G reactions. Lanes 2 and 6: DNase I cleavage control reactions. Lanes 3 and 7: DNase I cleavage reactions with $0.5 \mu\text{M}$ wt bZIP. Lanes 4 and 8: DNase I cleavage reactions with $3 \mu\text{M}$ wt bZIP. The positions of the core target sequences are indicated by the letter c; regions b and d are hypersensitivity cleavage regions. Region a contains additional footprints of the cognate AP-1 half-site and E-box half-site (CAC), as well as DNase I self-footprints and hypersensitivity cleavage regions. Region a displays the same pattern that previously published data exhibit and has been fully discussed in ref 9. Quantitative phosphorimaging analysis of the footprinting region is presented in Figure S2.

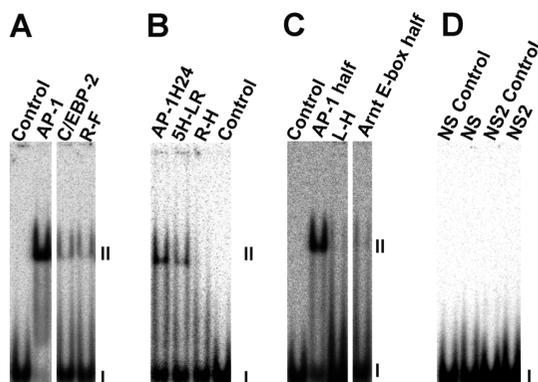


FIGURE 3: Qualitative EMSA analysis. Panels A–D represent different gels. (A) The C/EBP-2 and R-F sites with $1.8 \mu\text{M}$ s-wt bZIP; the AP-1 duplex serves as the specific control for these full sites. (B) The 5H-LR and R-H sites with $2.0 \mu\text{M}$ s-wt bZIP; the AP-1H24 site serves as the specific control for these half-sites. (C) The L-H site and the Arnt E-box half-site with $1.8 \mu\text{M}$ s-wt bZIP. The AP-1 half-site duplex serves as the specific control for these half-sites. (D) The nonspecific control duplexes, NS and NS2, with $2.0 \mu\text{M}$ s-wt bZIP. I denotes free DNA, and II denotes dimeric wt bZIP–DNA complexes.

We examined our EMSA conditions on nonspecific controls NS and NS2. No interactions were detected between NS or NS2 and the s-wt bZIP at a monomeric concentration of $2 \mu\text{M}$, which exceeds the highest s-wt bZIP concentration used in quantitative titrations (Figure 3; quantitative titrations were restricted to $\leq 1 \mu\text{M}$ s-wt bZIP, for higher concentrations occasionally lead to protein aggregation). Thus, our experimental conditions were restricted to selective wt bZIP–DNA interactions. Qualitative EMSA demonstrates that the e- and s-wt bZIP bind the C/EBP-2, R-F, and 5H-

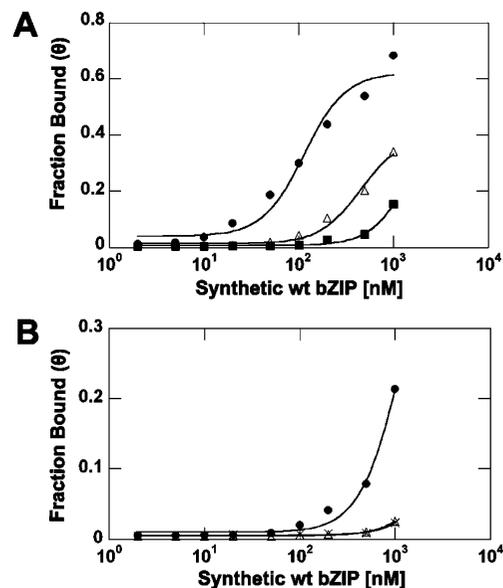


FIGURE 4: Representative equilibrium binding isotherms for the s-wt bZIP targeting (A) the C/EBP site (●), the Arnt E-box site (Δ), and the C/EBP-2 site (■) and (B) the 5H-LR site (●), the Arnt E-box half-site (Δ), and the L-H site (×). The latter two curves superimpose. Each isotherm was obtained from an individual EMSA titration.

LR sites, exhibit weak interaction with the L-H site and the Arnt E-box half-site (TCAC) (10), and exhibit no interaction with the R-H site (Figure 3; e-wt bZIP data not shown). All shifts in band migration correspond to the dimeric s-wt bZIP–DNA complex, as compared with the dimeric s-wt bZIP in complex with AP-1 site or the AP-1 half-site. Thus, the EMSA demonstrates that the wt bZIP selectively targets the C/EBP-2, R-F, 5H-LR, L-H sites and the Arnt-E-box half-sites.

For quantitative evaluation of wt bZIP–DNA binding affinities, apparent dimeric K_d values were obtained via thermodynamic EMSA titrations with the s-wt bZIP. Achieving completed equilibrium binding isotherms was restricted by the solubility of the s-wt bZIP; therefore, accurate K_d values could not be obtained in some cases. We therefore provide $\Delta\theta_{\text{app}}$ values, net bound DNA fractions at specified protein concentrations, as qualitative references of binding affinity.

The K_d values show that the affinity of the s-wt bZIP at the 5H-LR site is stronger than that at the L-H site; the clear difference in $\Delta\theta_{\text{app}}$ values confirms this qualitatively. Additionally, the affinity at the L-H site is higher than that at the R-H site, as the interaction of the s-wt bZIP with the L-H site is detected by EMSA, but that with the R-H site is not. Additionally, the K_d values show that the wt bZIP targets the C/EBP site with stronger affinity than the Arnt E-box or C/EBP-2 site, with qualitative confirmation from clear differences in $\Delta\theta_{\text{app}}$ values (Table 1 and Figure 4).

The wt bZIP Targets both the L and R Sequences. Our results show that the wt bZIP selectively targets the C/EBP, C/EBP-2, R-F, 5H-LR, L-H, and Arnt E-box sites. Which 4 bp sequences are contacted by the wt bZIP? To answer this question, all possible 4 bp sequences within these sites were extensively analyzed (Tables S1–S7; see Table 2 as a brief example). The results of these detailed analyses are summarized in Table 3. Accordingly, at the C/EBP-2 and R-F sites, the R sequence is the only plausible target for the wt

Table 2: Target Site Analysis of the 5H-LR Duplex

Possible Target Site ^a	Results ^c
5' -GAAA	Not a target site.
5' -AAAT	Not a target site.
5' -AATT	Not a target site.
5' -ATTG	Not a target site.
5' -TTGC (L)	Targeted subsite^b.
5' -TGCG (R)	Targeted subsite^b.
5' -GCGT	Not a target site.
5' -CGTT	Not a target site.
5' -GTTT	Not a target site.
5' -TTTG	Not a target site.
5' -TTGA	Not a target site.
TCGAGGAAAT TTGCG TTTGAAGGTT	
ACGTCCTTT AACG CAACTTCCAA	
AACT-5'	Not a target site.
AAAC-5'	Not a target site.
CAAA-5'	Not a target site.
GCAA-5'	Not a target site.
CGCA-5'	Not a target site.
ACGC-5'	Not a target site.
AACG-5'	Not a target site.
TAAAC-5'	Not a target site.
TTAA-5'	Not a target site.
TTTA-5'	Not a target site.
CTTT-5'	Not a target site.

^a Core target sequences are in bold, and the entire inserted sequences between flanking sequences are underlined. ^b TTGC is the L subsite, and TGCG is the R subsite. ^c See Table S4 for detailed analysis.

Table 3: Summary of Target Site Analyses

DNA Site	Half Sites/Subsites ^a		Locations ^b
	Coding Strand	Complementary Strand	
C/EBP	L, R (TTGCG)	L, R (TTGCG)	TTGCG CAA AAC GCGT
C/EBP-2	R	R	- TGCG CA- -AC GCGT -
R-F	R	R	TGCG CGCA ACGC GCGT
Arnt E-box	TCAC	TCAC	TCAC GTGA AGT GC ACT
5H-LR	L, R (TTGCG)		TTGCG --- AACGC---
L-H	L		TTGC ---- AACG----
R-H	R		- TGCG --- -ACGC---

^a The basic region of the wt bZIP targets only the Arnt E-box half-site (TCAC) and the TTGC (L) and TGCG (R) subsites. Overlapping L and R subsites result in the overall 5 bp TTGCG half-site (5H-LR). ^b Locations of half-sites and subsites within core target sites. Half-sites and subsites are in bold, boxed in gray. See detailed analyses in Tables S1–S7.

bZIP. Thus, although half-site binding was not detected at the R-H site, which is a single R sequence, full-site binding by the wt bZIP at the C/EBP-2 and R-F sites demonstrates that the wt bZIP selectively targets the R sequence.

According to our analysis, the wt bZIP can target only L and R sequences within the C/EBP site; no other sequence is a plausible target (Table 3). At the C/EBP site, does the wt bZIP contact both L and R sequences? To answer this question, we investigated the wt bZIP targeting the C/EBP-2 and Arnt E-box sites. At the C/EBP-2 site, the wt bZIP targets only the overlapping R sequences, positioned identically as in the C/EBP site. Thus, the wt bZIP–C/EBP-2

complex demonstrates how the wt bZIP targets only the overlapping R sequences in the C/EBP site without any influence from the L sequence. Furthermore, the wt bZIP demonstrates the same binding affinities and $\Delta\theta_{app}$ values at the L sequence (L-H site) and Arnt E-box half-site, and therefore, we consider wt bZIP interactions at these two sites to be thermodynamically equivalent (Figure 4 and Table 1). Thus, the wt bZIP–Arnt E-box complex represents how the wt bZIP dimer targets the abutting L sequences in the C/EBP site without any influence from the R sequence. Thus, we examined C/EBP-2 and the Arnt E-box sites because the C/EBP site, TTGCGCAA, embeds both L and R sequences, and we cannot dissect their individual contributions to overall binding (Table 3).

Although the wt bZIP exhibits a stronger affinity for the Arnt E-box site than for the C/EBP-2 site, neither case individually accounts for the significant binding affinity measured at the C/EBP site: the affinity at this site is substantially stronger than those at the Arnt E-box and C/EBP-2 sites by 20- and 130-fold, respectively, with clear differences in $\Delta\theta_{app}$ values (Table 1 and Figure 4). As our results show that the wt bZIP can selectively target both the L and R sequences in the C/EBP site, this analysis suggests that both interactions contribute to binding affinity at the C/EBP site.

The wt bZIP Targets Overlapping L and R Subsites within the TTGCG Half-Site, Resulting in Increased Binding Affinities. We therefore used the 5H-LR site (TTGCG) to test our findings: the wt bZIP interacts with both L and R sequences, and both interactions contribute to its affinity for the C/EBP site. Only overlapping L and R sequences on the coding strand of the 5H-LR site are targeted (Table 3). The wt bZIP targets the 5H-LR site with ≥ 10 -fold stronger affinity and a clearly higher $\Delta\theta_{app}$ value compared to those of the individual L or R sequences [L-H or R-H sites, respectively (Figure 4 and Table 1)]: localizing the wt bZIP at the individual L or R sequence does not explain the high affinity measured at the 5H-LR site. Despite the fact that the wt bZIP–L interaction is stronger than the wt bZIP–R interaction, our analysis again demonstrates that both interactions can occur at the C/EBP site. Consequently, we conclude that interactions at the L and R sequences contribute to the overall affinity between the wt bZIP and the 5H-LR site.

The wt bZIP targets the C/EBP site with an affinity 70-fold stronger than that of the 5H-LR site, with clear differences in $\Delta\theta_{app}$ values (Figure 4 and Table 1). This result is consistent with increases between the affinities of half- and full-site complexes of the wt bZIP with AP-1 and Arnt E-box (44- and 30-fold, respectively) (10) and the GCN4 bZIP with AP-1 and CRE (30- and 80-fold, respectively) (14). Thus, the L and R subsites constitute the effective half-site sequence TTGCG in the 5H-LR site selectively targeted by the wt bZIP when binding to the C/EBP site, and the significant affinity at the 5H-LR site contributes to the strong affinity measured at the C/EBP site.

DISCUSSION

Our results demonstrate that the wt bZIP contacts both 4 bp L and R subsites within the 5H-LR site. At the 5H-LR site, how does the wt bZIP contact two overlapping subsites on the same DNA strand? We propose that the wt bZIP

Table 4: Numbering of Target Sites

Target Site	Sequence							
AP-1	T ₄	G ₃	A ₂	C ₁	T ₀	C	A	
	A _{4'}	C _{3'}	T _{2'}	G _{1'}	A _{0'}	G	T	
CRE	T ₄	G ₃	A ₂	C ₁	G ₀	T	C	A
	A _{4'}	C _{3'}	T _{2'}	G _{1'}	C _{0'}	A	G	T
5H-LR	T ₄	T ₃	G ₂	C ₁	G ₀			
	A _{4'}	A _{3'}	C _{2'}	G _{1'}	C _{0'}			

continually slides between the two subsites; such motion likely requires constant conformational changes in the basic regions.

Crystal structures of the GCN4 bZIP bound to the AP-1 or CRE sites demonstrate how each basic region contacts a 4 bp TGAC half-site in the DNA major groove (3–5). As the wt bZIP comprises the same GCN4 basic region used in the crystallographic studies, we expect the same interactions between wt bZIP and DNA. These crystal structures show that Asn235 forms hydrogen bonds with T4 and C3' (Table 4). Similarly at the L subsite in 5H-LR, Asn235 of the wt bZIP can form H-bonds with T4 and A3'. At the R subsite in 5H-LR, Asn235 can form H-bonds with T3 and C2', as in the GCN4 complexes with the AP-1 and CRE sites. Therefore, Asn235 can slide between the T4 and T3 bases and A3' and C2' bases in the 5H-LR site; such sliding increases the probability of the formation of H-bonds between Asn235 and specific DNA bases in the major groove.

In the asymmetric AP-1 complex, one GCN4 bZIP basic region uses Arg243 to make bidentate H-bonds to N7 and O6 of G1', and the other basic region donates H-bonds to the phosphodiester backbone (Table 4) (3). In the symmetric CRE complex, the GCN4 bZIP makes a hybrid of the interactions made to AP-1: Arg243 forms a single H-bond to G1' and direct and water-mediated interactions with the phosphodiester backbone (4, 5). At the L subsite in 5H-LR, Arg243 of the wt bZIP can form the same contacts with G1' and the phosphodiester backbone as in the GCN4 complexes with the AP-1 or CRE sites. At the R subsite in 5H-LR, Arg243 can make nonspecific contacts with the phosphodiester backbone near C0', as no specific interaction with cytosine is likely. Therefore, like Asn235 discussed above, Arg243 can slide between the G1' and C0' bases in the 5H-LR site to increase the probability of the formation of specific H-bonds with the DNA target.

This analysis explains how the wt bZIP accesses both the L and R subsites in 5H-LR: Asn235 can slide between T3 and T4 and between A3' and C2' at the 5'-end of the 5H-LR site, and Arg243 can slide between G1' and C0' at the 3'-end (Table 4). Therefore, the entire basic region can slide between the L and R subsites to maximize contacts at the protein–DNA interface (Figure 5). Such sliding increases the probability of forming hydrogen bonds, thereby increasing the macroscopic K_{on} and binding affinity at the 5H-LR site, as observed.

Such a mechanism for sliding between the L and R subsites requires conformational changes in the wt bZIP basic regions (and possibly DNA, as well), in particular, for residues making specific contacts to DNA, including Asn235 and Arg243. Generally, DNA binding induces α -helical structure in the bZIP basic region, thereby maximizing specific DNA contacts, as seen in the “induced fit” concept describing

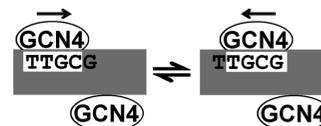


FIGURE 5: The wt bZIP targets the 5H-LR site. The wt bZIP dimer (depicted as a pair of ovals of the GCN4 basic regions) uses one basic region to interact selectively with the overall 5 bp half-site, while the other basic region binds DNA nonspecifically. The basic region constantly slides between two overlapping 4 bp subsites.

enzyme–ligand association. Thus, the basic region of the wt bZIP can exist in more than one conformation as it slides between the L and R subsites. Due to multiple conformations, the conformational entropy of the bound state will increase and the free energy of the wt bZIP–5H-LR complex will decrease to further stabilize the complex, as compared to the wt bZIP existing in a restricted number of bound states in complex with each individual subsite.

This binding mechanism is supported by a one-dimensional sliding mechanism used by proteins to facilitate the rapid search for specific DNA sequences (15, 16). Gorman et al. used total internal reflection fluorescence microscopy to show that the Msh2–Msh6 protein dimer can slide along DNA at a rate approaching 800 bp/s (17). Winter et al. used quantitative filter binding assays to estimate a 100 bp sliding length for the Lac repressor before dissociation from DNA under physiological conditions (18). Thus, at the protein–DNA interface in the wt bZIP–5H-LR complex, the bZIP basic regions can slide rapidly between two overlapping subsites to maximize binding affinity.

Solution NMR of the GCN4 bZIP shows that the leucine zipper is a stable and helical coiled coil while the basic region is substantially helical but highly dynamic (19, 20). Columbus and Hubbell performed solution EPR on the free GCN4 bZIP α -helix and its complex with the AP-1 site (21). When the GCN4 bZIP binds to AP-1, backbone motions in the basic region are dampened, but a gradient of mobility persists; this indicates significant internal flexibility within the basic region. Such flexibility can allow the requisite conformational changes in the wt bZIP basic regions for sliding between the L and R subsites and maximizing interactions.

In this work, we propose a binding mechanism for proteins in complex with DNA: a protein can target multiple subsites within a DNA target site by sliding between subsites with conformational changes occurring at the protein–DNA interface. Via this binding mechanism, a protein realizes high DNA-binding affinity by including interactions with multiple target subsites on the same DNA strand. This mechanism also leads to high sequence selectivity, as DNA binding becomes more restrictive due to recognition of a larger DNA target sequence. This mechanism may also explain why some transcription factors in complex with gene regulatory sequences have not been amenable to characterization by high-resolution solution studies. Nature may harness multiple strategies with the goal of high affinity and highly specific protein–DNA recognition. This binding mechanism expands our understanding of what constitutes the DNA target site in protein–DNA interactions.

ACKNOWLEDGMENT

We thank Alevtina Pavlenko for technical assistance and the reviewer who improved our Discussion with suggestions about conformational changes in the basic regions.

SUPPORTING INFORMATION AVAILABLE

Target site analyses. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

1. Struhl, K. (1989) Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eucaryotic transcriptional regulatory proteins. *Trends Biochem. Sci.* *14*, 137–140.
2. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* *240*, 1759–1764.
3. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein-DNA complex. *Cell* *71*, 1223–1237.
4. König, P., and Richmond, T. J. (1993) The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* *233*, 139–154.
5. Keller, W., König, P., and Richmond, T. J. (1995) Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *J. Mol. Biol.* *254*, 657–667.
6. Lajmi, A. R., Wallace, T. R., and Shin, J. A. (2000) Short, hydrophobic, alanine-based proteins based on the bZIP motif: overcoming inclusion body formation and protein aggregation during overexpression, purification, and renaturation. *Protein Expression Purif.* *18*, 394–403.
7. Lajmi, A. R., Lovrencic, M. E., Wallace, T. R., Thomlinson, R. R., and Shin, J. A. (2000) Minimalist, alanine-based, helical protein dimers bind to specific DNA sites. *J. Am. Chem. Soc.* *122*, 5638–5639.
8. Bird, G. H., Lajmi, A. R., and Shin, J. A. (2002) Sequence-specific recognition of DNA by hydrophobic, alanine-scanning mutants of the bZIP motif investigated by fluorescence anisotropy. *Biopolymers* *65*, 10–20.
9. Fedorova, A. V., Chan, I-S., and Shin, J. A. (2006) The GCN4 bZIP can bind to noncognate gene regulatory sequences. *Biochim. Biophys. Acta* *1764*, 1252–1259.
10. Chan, I-S., Fedorova, A. V., and Shin, J. A. (2007) The GCN4 bZIP targets noncognate gene regulatory sequences: quantitative investigation of binding at full and half sites. *Biochemistry* *46*, 1663–1671.
11. Bird, G. H., and Shin, J. A. (2002) MALDI-TOF mass spectrometry characterization of hydrophobic basic region/leucine zipper proteins. *Biochim. Biophys. Acta* *1597*, 252–259.
12. Xie, Y., and Wetlaufer, D. B. (1996) Control of aggregation in protein refolding: the temperature-leap tactic. *Protein Sci.* *5*, 517–523.
13. Metallo, S. J., and Schepartz, A. (1994) Distribution of labor among bZIP segments in the control of DNA affinity and specificity. *Chem. Biol.* *1*, 143–151.
14. Hollenbeck, J. J., and Oakley, M. G. (2000) GCN4 binds with high affinity to DNA sequences containing a single consensus half-site. *Biochemistry* *39*, 6380–6389.
15. von Hippel, P. H., and Berg, O. G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.* *264*, 675–678.
16. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* *32*, 3040–3052.
17. Gorman, J., Chowdhury, A., Surtees, J. A., Shimada, J., Reichman, D. R., Alani, E., and Greene, E. C. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell* *28*, 359–370.
18. Winter, R. B., Berg, O. G., and von Hippel, P. H. (1981) Diffusion-driven mechanism of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: Kinetic measurements and conclusions. *Biochemistry* *20*, 6961–6977.
19. Saudek, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R., and Pastore, A. (1991) Solution structure of the basic region from the transcriptional activator GCN4. *Biochemistry* *30*, 1310–1317.
20. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G., III (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol.* *285*, 2133–2146.
21. Columbus, L., and Hubbell, W. L. (2004) Mapping backbone dynamics in solution with site-directed spin labeling: GCN4–58 bZip free and bound to DNA. *Biochemistry* *43*, 7273–7287.
22. Hill, D. E., Hope, I. A., Macke, J. P., and Struhl, K. (1986) Saturation mutagenesis of the yeast *his3* regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* *234*, 451–457.
23. Pu, W. T., and Struhl, K. (1991) Highly conserved residues in the bZIP domain of yeast GCN4 are not essential for DNA binding. *Mol. Cell. Biol.* *11*, 4918–4926.

BI800355T