The bZIP Dimer Localizes at DNA Full-Sites Where Each Basic Region Can Alternately Translocate and Bind to Subsites at the Half-Site

I-San Chan,† Taufik Al-Sarraj,† S. Hesam Shahrvan,‡ Anna V. Fedorova,‡ and Jumi A. Shin*†‡

†Department of Chemistry, University of Toronto, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada
‡Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College Street, Toronto, Ontario M5S 3G9, Canada

ABSTRACT: Crystal structures of the GCN4 bZIP (basic region/leucine zipper) with the AP-1 or CRE site show how each GCN4 basic region binds to a 4 bp cognate half-site as a single DNA target; however, this may not always fully describe how bZIP proteins interact with their target sites. Previously, we showed that the GCN4 basic region interacts with all 5 bp in half-site TTGCG (termed SH-LR) and that SH-LR comprises two 4 bp subsites, TTGC and TGCG, which individually are also target sites of the basic region. In this work, we explore how the basic region interacts with SH-LR when the bZIP dimer localizes to full-sites. Using AMBER molecular modeling, we simulated GCN4 bZIP complexes with full-sites containing SH-LR to investigate in silico the interface between the basic region and SH-LR. We also performed in vitro investigation of bZIP–DNA interactions at a number of full-sites that contain SH-LR versus either subsite: we analyzed results from DNase I footprinting and electrophoretic mobility shift assay (EMSA) and from EMSA titrations to quantify binding affinities. Our computational and experimental results together support a highly dynamic DNA-binding model: when a bZIP dimer localizes to its target full-site, the basic region can alternately recognize either subsite as a distinct target at SH-LR and translocate between the subsites, potentially by sliding and hopping. This model provides added insights into how α-helical DNA-binding domains of transcription factors can localize to their gene regulatory sequences in vivo.

T ranscription factors use their DNA-binding domains to search for and localize to their cognate gene regulatory sequences to govern gene expression. During the search phase, these domains translocate along genomic DNA, and the protein–DNA interactions are mostly nonspecific. During the localization phase, however, the same domains contact their cognate target sites, and the interactions are sequence-specific. The processes of search and localization, as well as transition between them, are not well-understood.1 We therefore explored sequence-selective (subspecific) interactions between DNA-binding domains and noncognate target sites. Research on such protein–DNA interactions can provide an improved understanding of how the same DNA-binding domains execute the coupled search and localization tasks in vivo. At these regulatory sequences, DNA-binding domains of basic region/leucine zipper (bZIP) transcription factors, such as yeast GCN4, bind to DNA as a dimer of short, continuous α-helices (~60 residues).2 The bZIP motif is the simplest structure used by transcription factors to contact specific DNA sequences. Therefore, a thorough understanding of the mechanisms that such a simple structure can use to interact with DNA target sites will facilitate research on artificial transcription factors and more complicated DNA-binding proteins.

Each bZIP α-helix comprises a C-terminal leucine zipper for dimerization and N-terminal basic region for DNA binding (for a review, see ref 2). McKnight and co-workers exchanged leucine zippers between bZIP proteins GCN4 and C/EBP (CCAAT/enhancer binding protein) and showed that sequence selectivity for DNA binding follows the basic regions.3 The GCN4 bZIP dimers cognate full-site AP-1 (7 bp, TGA(C/G)TCA), while the basic region binds to 4 bp cognate half-site TGA(C/G).4 The GCN4 bZIP also targets the CRE site (8 bp, TGACGTCA), which differs from AP-1 by one central base pair.5 The GCN4 basic region and can overlap to give 5 bp TTGCG.7,8 Therefore, we named 4 bp TTGC and TGCG as "R" (right), the 5′ and 3′ subsites of 5-bp TTGCG, respectively; we termed this 5 bp sequence “SH-LR” as a 5 bp hybrid of subsites L and R. Additionally, the basic region exhibited ≥10-fold higher affinity at SH-LR than that at either L.

Received: May 31, 2012
Revised: August 1, 2012
Published: August 2, 2012
or R. This indicates that the basic region interacts with 5 bp TTGCG at 5H-LR, not just with L or R. Hence, 5H-LR acts as an overall half-site. Not only is 5H-LR 1 bp longer than the 4 bp cognate half-site, but it contains two 4 bp subsites that are also target sites of the same basic region. We hence surmised that the basic region may not recognize 5H-LR as a single DNA target.

When the bZIP dimer localizes to full-sites, how does the basic region interact with 5H-LR? We explored the answer in this work. First, we examined whether the basic region recognizes 5H-LR always as a single target or recognizes either subsite as part of its interaction with 5H-LR. If the latter is possible, how would it be achieved? We examined the following possibilities. Would the basic region bind to either subsite followed by complete dissociation from DNA, resulting in a mixture of some GCN4 basic regions specifically contacting subsite L and others contacting subsite R? Would the basic region be mobile along 5H-LR and thus recognize the subsites alternately?

If the second situation is possible, the basic region may not recognize 5H-LR as a single DNA target. When the bZIP dimer localizes to full-sites, how does the basic region interact with 5H-LR? We explored the answer in this work. First, we examined whether the basic region recognizes 5H-LR always as a single target or recognizes either subsite as part of its interaction with 5H-LR. If the latter is possible, how would it be achieved? We examined the following possibilities. Would the basic region bind to either subsite followed by complete dissociation from DNA, resulting in a mixture of some GCN4 basic regions specifically contacting subsite L and others contacting subsite R? Would the basic region be mobile along 5H-LR and thus recognize the subsites alternately?

If the second situation is possible, the basic region must translocate between subsites, given that their positions differ by 1 bp. Solution electron paramagnetic resonance (EPR) studies show that the GCN4 basic region exhibits backbone mobility even when bound to the cognate AP-1 site; this suggests the possibility that the basic region can be mobile along SH-LR. DNA-binding proteins also exhibit mobility by translocating along genomic DNA during target-site search. There are four mechanisms for rapid protein translocation along DNA: (i) sliding, diffusion along DNA without dissociation; (ii) hopping, dissociation and reassociation between closely spaced DNA segments; (iii) jumping, dissociation from DNA and rebinding to a distant DNA segment; and (iv) intersegment transfer, moving between two segments brought close by looped DNA. These mechanisms have been exhibited by various proteins and captured in vitro by a variety of techniques, e.g., sliding of protein β clamp by fluorescence resonance energy transfer (FRET) and hopping of human DNA repair factor RAD54 by atomic force microscopy (AFM). Sliding and hopping are relevant to closely spaced DNA segments. Therefore, we examined the possibility that the basic region slides or hops to translocate between 5H-LR’s subsites.

We performed in silico and in vitro studies of how the basic region interacts with 5H-LR. The native GCN4 bZIP used for in silico studies and chimeric wt bZIP used for in vitro studies both contain the GCN4 basic region but dimerize via the GCN4 or C/EBP leucine zipper, respectively (Figure 1). We previously generated this chimeric bZIP protein and termed it wt bZIP (“wild type”), which contains the wild-type GCN4 basic region, to compare with our engineered mutant proteins. We showed that wt bZIP exhibits the function and structure of the GCN4 bZIP, regardless of zippers; therefore, we expected the GCN4 bZIP and wt bZIP to make identical interactions with 5H-LR. The in silico studies included AMBER (Assisted Model Building with Energy Refinement) simulations to yield snapshots of the interface between the basic region and SH-LR to analyze DNA sequences recognized by the basic region. The in vitro studies were used to examine DNA-binding affinities of wt bZIP at full-site C/EBP (eponymous cognate

### Figure 1. Protein sequences. Basic regions are shown in bold with leucine zippers underlined. (A) Native GCN4 bZIP (residues 226–281). (B) wt bZIP: the GCN4 basic region (residues 226–252) fused to C/EBP leucine zipper (residues 310–338). The N-terminal Met of e-wt bZIP was removed during post-translational modification.

### Figure 2. DNA sequences. Core target sequences are shown in bold with inserted sequences underlined. (A) Sequences used in DNase I footprinting analysis. Only the core target site and surrounding flanking sequences from each DNA duplex are shown. The flanking sequences of AP-1 are from base pair -87 to -102 of the yeast his3 promoter region, those of other target sites are identical to those for the EMSA. AP-1 serves as a control. (B) Sequences used in EMSA (24 bp). Flanking sequences, identical for all DNA duplexes, were chosen to minimize DNA secondary structure. For the same reason, core target sites in the AP-1H24 and SH-LR duplexes were shifted by 1 bp toward the 3’ end. Other duplexes are listed in Figure S3 of the Supporting Information. AP-1, CRE, and AP-1H24 serve as controls. (C) Sequences used in molecular modeling. AP-1 and CRE are from the GCN4–DNA crystal structures. C/EBP and C/EBP-1 are same as those for the EMSA.

### Table 1

<table>
<thead>
<tr>
<th>Sequences used in footprinting analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP-1</td>
</tr>
<tr>
<td>C/EBP-1</td>
</tr>
<tr>
<td>AC</td>
</tr>
<tr>
<td>AC-1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequences used in EMSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP-1</td>
</tr>
<tr>
<td>CRE</td>
</tr>
<tr>
<td>XRE1</td>
</tr>
<tr>
<td>AC</td>
</tr>
<tr>
<td>C/EBP</td>
</tr>
<tr>
<td>C/EBP-1</td>
</tr>
<tr>
<td>C/EBP-2</td>
</tr>
<tr>
<td>AmiE-box</td>
</tr>
<tr>
<td>SH-LR</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Sequences used in molecular modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP-1</td>
</tr>
<tr>
<td>C/EBP</td>
</tr>
<tr>
<td>CRE</td>
</tr>
<tr>
<td>XRE1</td>
</tr>
<tr>
<td>AC</td>
</tr>
<tr>
<td>C/EBP</td>
</tr>
<tr>
<td>C/EBP-1</td>
</tr>
<tr>
<td>C/EBP-2</td>
</tr>
<tr>
<td>AmiE-box</td>
</tr>
<tr>
<td>SH-LR</td>
</tr>
</tbody>
</table>

### Figure 1.

**Protein sequences.** Basic regions are shown in bold with leucine zippers underlined. (A) Native GCN4 bZIP (residues 226–281). (B) wt bZIP: the GCN4 basic region (residues 226–252) fused to C/EBP leucine zipper (residues 310–338). The N-terminal Met of e-wt bZIP was removed during post-translational modification.

### Figure 2.

**DNA sequences.** Core target sequences are shown in bold with inserted sequences underlined. (A) Sequences used in DNase I footprinting analysis. Only the core target site and surrounding flanking sequences from each DNA duplex are shown. The flanking sequences of AP-1 are from base pair -87 to -102 of the yeast his3 promoter region, those of other target sites are identical to those for the EMSA. AP-1 serves as a control. (B) Sequences used in EMSA (24 bp). Flanking sequences, identical for all DNA duplexes, were chosen to minimize DNA secondary structure. For the same reason, core target sites in the AP-1H24 and SH-LR duplexes were shifted by 1 bp toward the 3’ end. Other duplexes are listed in Figure S3 of the Supporting Information. AP-1, CRE, and AP-1H24 serve as controls. (C) Sequences used in molecular modeling. AP-1 and CRE are from the GCN4–DNA crystal structures. C/EBP and C/EBP-1 are same as those for the EMSA.
target of the C/EBP bZIP; this site comprises two copies of SH-LR), half-site SH-LR, subsites L and R, and derivative target sites (Figure 2); we used DNase I footprinting and electrophoretic mobility shift assay (EMSA) to evaluate sequence-selective DNA binding, and EMSA titrations to determine $K_d$ values of bZIP–DNA complexes. These data were analyzed to explore how the basic region interacts with subsites in SH-LR and the possibility of basic region sliding and hopping. Our results provide added insights into how $\alpha$-helical proteins localize to their target sites and transit between target-site search and localization in vivo.

## EXPERIMENTAL PROCEDURES

### In Silico Studies via Molecular Modeling.

We used AMBER software (version 9) and force field f99SB and performed energy minimization to obtain snapshots of GCN4 bZIP complexes with the C/EBP or C/EBP-1 site. The force field was chosen because of its improved protein backbone parameters suitable for simulations of $\alpha$-helical proteins.16 For every simulation, a truncated octahedral unit cell, with an 8.00 Å buffer between the solute and box edge, was solvated with explicit TIP3P water. Na$^+$ ions were added to achieve electroneutrality. The simulations were conducted in 20000 steps, performed using a 9.00 Å cutoff on real-space interactions, and run on four nodes at the Center for Molecular Design and Preformulations (CMDP), University Health Network (Toronto, ON), on a SGI Onyx 3800 supercomputing system. The distances were calculated using Coot (Crystallographic Object-Oriented Toolkit).18 The images of the energy-minimized complexes were documented using VMD (Visual Molecular Dynamics).19 All atom locants were documented using IUPAC nomenclature (Greek superscript letters), after transcription from the Protein Data Bank (PDB) and modeling files (uppercase Roman letters), e.g., $N^{\alpha}$ = ND2. See section S1 of the Supporting Information for additional details of the simulations.

For our control experiments, 1YSA$^4$ and 1DGC$^5$ (crystal structures of the GCN4 bZIP with the AP-1 and CRE sites, respectively) from the PDB were simulated to yield energy-minimized complexes 1YSA$^{em}$ and 1DGC$^{em}$, respectively. 1YSA and 1DGC both contain the GCN4 bZIP, residues 226–281 (Figure 1A), with N-terminal MK residues in 1YSA only. See Figure 2 for DNA sequences. See section S1.2 of the Supporting Information. To create $\alpha_1$ and $\beta_1$, initial structures of the GCN4 bZIP with the C/EBP and C/EBP-1 sites, respectively, we used the “simple mutate” function in Coot to modify the bases of the AP-1 duplex in energy-minimized 1YSA$^{em}$ to match those of the C/EBP and C/EBP-1 duplexes (Figure 2C); the remainder of 1YSA$^{em}$ was unchanged. Initial structures $\alpha_1$ and $\beta_1$ were simulated to yield energy-minimized $\alpha_1^{em}$ and $\beta_1^{em}$, the first snapshots of the GCN4 bZIP complexes with C/EBP and C/EBP-1, respectively.

In addition to this “direct” approach to obtaining snapshots $\alpha_1^{em}$ and $\beta_1^{em}$ we pursued a “reverse” approach to obtain additional snapshots of the same complexes. We utilized $\alpha_1^{em}$ and restrained four distances using the “makeDIST_RST” function in AMBER: distances from $N^{\alpha}_2$ of Asn235 in BRA and BrR (basic regions in the left and right halves, respectively, of the bZIP–DNA complex) to O4 of the T4 and T19 bases, respectively, were restrained to 2.30–2.80 Å, while distances from $N^{\alpha}_2$ of Asn235 in BRa and BrR to O4 of the T4 and T19 bases, respectively, were restrained to 5.00–5.50 Å (Table S2 of the Supporting Information). These distance settings were based on those from $N^{\alpha}_2$ of Asn235 in BRa to O4 of the T4 and T19 bases in the left half of $\alpha_1^{em}$. This simulation was run for 20000 steps, without periodic boundary conditions and with a distance-dependent dielectric (option eedmeth = 5), to create initial structure $\alpha_2$ (with C/EBP). These distance restraints made $\alpha_2$ different from $\alpha_1$ (see the analysis and comparison of initial structures $\alpha_1$ and $\alpha_2$ in section S1.1 of the Supporting Information). To create initial structure $\beta_2$ (with C/EBP-1), we modified DNA bases 11 and 18 of $\alpha_2$ (Figure 2C); $\alpha_2$ and $\beta_2$ were otherwise identical. We simulated $\alpha_2$ and $\beta_2$ to obtain energy-minimized $\alpha_2^{em}$ and $\beta_2^{em}$ (second snapshots of the complexes with C/EBP and C/EBP-1, respectively), using the same conditions that were used to give 1YSA$^{em}$, 1DGC$^{em}$ $\alpha_1^{em}$ and $\beta_1^{em}$. Unlike the $\alpha_1$ simulation, these simulations were performed without any distance restraints.
In Vitro Studies. Protocols for our in vitro studies have been previously published. See ref 20 for production of e-wt bZIP. See ref 21 for purification of e- and s-wt bZIP by reversed-phase HPLC and verification by ESI-MS. See ref 21 for protocols of DNase I footprinting analysis and EMSA. See ref 8 for a determination of dimeric $K_d$ values. Section S5 of the Supporting Information provides a brief summary of these protocols.

RESULTS AND DISCUSSION

In Silico Studies via Molecular Modeling of GCN4–DNA Complexes. To examine how the GCN4 basic region interacts with the SH-LR half-site, we first examined whether the basic region recognizes SH-LR solely as a single DNA target or recognizes either subsite L or R as part of its interaction with SH-LR. We therefore explored the interface between the basic region and SH-LR by analyzing snapshots of the interface from the simulated GCN4 bZIP complexes with the C/EBP or C/EBP-1 site.

We generated these complexes via AMBER energy minimization, using initial structures based on the crystal structure of the GCN4 bZIP with the AP-1 site (1YSA$^4$). The C/EBP site comprises two SH-LR half-sites (Figure 2). A single base-pair mutation at the 3' end of the C/EBP site gives C/EBP-1; thus, the C/EBP and C/EBP-1 duplexes differ only at bases 11 and 18 (Figure 2C). As a result, the C/EBP-1 site comprises one SH-LR and one R sequence, which are in the same positions as they are in the C/EBP site. We performed energy minimization on crystal structures 1YSA$^4$ and 1DGC$^5$ (the GCN4 bZIP with the CRE site) to obtain 1YSA$^6_{em}$ and 1DGC$^6_{em}$, respectively, for control experiments (see section S1.2 of the Supporting Information for an examination of control experiments). We found that the same nine basic-region residues made direct H-bonds to each base pair in 1YSA$^6_{em}$ but also in the GCN4–DNA complexes: Arg232, Arg234, Asn235, Thr236, Arg240, Arg241, Ser242, and Arg243 to indicate the ends of the recognized DNA sequence (although the complexes were simulated with TIP3P water). We also did not include van der Waals interactions, because they are not helpful in distinguishing the ends of sequences recognized by the basic region in the GCN4–DNA crystal structures.$^4$–$^6$

Not only did we explore where H-bonds occur in snapshots $\alpha_1$–$\alpha_2$, $\beta_1$–$\beta_2$, and $\beta_2$–$\beta_2$em, but as described in later discussion, we also compared distances between the same H-bonding pairs among these four snapshots to explore atomic displacements, which can suggest mobility of these atoms at the GCN4–DNA interfaces. Therefore, rather than detecting H-bonds directly, we used the <3.00 Å distance between a H-bond donor and acceptor to indicate a direct H-bond.

We found that the same nine basic-region residues made direct H-bonds with DNA not only in snapshots $\alpha_1$–$\alpha_2$, $\beta_1$–$\beta_2$, and $\beta_2$–$\beta_2$em, but also in the GCN4–DNA crystal structures: Arg232, Arg234, Asn235, Thr236, Arg240, Arg241, Ser242, Arg243, and Arg245 (Tables S5 and S6 of the Supporting Information). This finding indicates that the basic region uses the same residues to interact with SH-LR and the cognate half-sites.

In the following section, we analyzed the DNA sequence recognized by each basic region in snapshots $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$em. We used direct H-bonds donated by N$^\delta$ of Asn235 and N$^\rho_2$ of Arg243 to indicate the ends of the recognized DNA sequences. We used these two H-bond donors because Asn235 and Arg243 are the only residues that make direct H-bonds to the DNA bases; all other DNA-binding residues made direct H-bonds to the DNA backbone only (Tables S5 and S6 of the Supporting Information). Also, the same H-bond donors were used in the above discussion to indicate that each basic region recognizes the 4 bp cognate half-site, in the GCN4–DNA crystal structures [1DGC (Figure 3), 2DGC, and 1YSA$^4$]–$^6$. In this way, we examined whether the basic region recognizes SH-LR solely as a single DNA target or recognizes either subsite as part of its interaction with SH-LR.

DNA Sequences Recognized in Snapshots of the GCN4 bZIP Complex with the C/EBP or C/EBP-1 Site. We examined snapshots $\alpha_1$–$\alpha_2$ and $\beta_1$–$\beta_2$ of the complex with C/EBP, whose left and right halves each contain SH-LR. These
snapshots differed in their initial structures: BR<sub>α</sub> recognized 4 bp sub site R in α<sub>1</sub> versus 5 bp SH-LR in α<sub>1</sub> (see analysis of α<sub>1</sub> vs α<sub>2</sub> in section S1.1 of the Supporting Information). In the left halves of both snapshots, BR<sub>α</sub> recognized 4 bp sub site L but not 5 bp SH-LR, as indicated by H-bonds made by N<sup>60</sup> of Asn235 and N<sup>60</sup> of Arg243 (Figure 3). In the right halves of both snapshots, N<sup>60</sup> of Arg243 from BR<sub>δ</sub> contacted the 3′ end of SH-LR when N<sup>60</sup> of Asn235 was situated between the two 5′ end base pairs, but N<sup>60</sup> of Asn235 did not reach the 5′ end of SH-LR. We found that no basic region recognized the full length of SH-LR in snapshots α<sub>1</sub> em and α<sub>2</sub> em.

We also examined snapshots β<sub>1</sub> em and β<sub>2</sub> em of the complex with C/EBP-1. The left half of each snapshot contained SH-LR, but BR<sub>α</sub> recognized only 4 bp sub site L (Figure 3). Thus, we found that BR<sub>α</sub> did not recognize the full length of SH-LR in snapshots β<sub>1</sub> em and β<sub>2</sub> em.

These results from snapshots α<sub>1</sub> em, α<sub>2</sub> em, β<sub>1</sub> em and β<sub>2</sub> em suggest that the basic region can recognize 4 bp sub site L as a distinct target at SH-LR. Moreover, the right half of the complex with C/EBP-1 contained one R sequence. BR<sub>α</sub> recognized this R sequence in snapshot β<sub>1</sub> em (Figure 3); also, N<sup>60</sup> of Asn235 and N<sup>60</sup> of Arg243 from BR<sub>δ</sub> in snapshots β<sub>1</sub> em and β<sub>2</sub> em exhibited a very similar H-bonding pattern. Beyond this R sequence, BR<sub>α</sub> established direct H-bonds only to the DNA backbone, but not to DNA bases. These results suggest that the basic region can recognize 4 bp sub site R as a distinct target, regardless of neighboring base pairs, at SH-LR. These analyses together suggest that the basic region does not recognize SH-LR solely as a single target; these analyses offer no evidence of recognition of the full length of SH-LR (i.e., recognition of both subsites simultaneously). However, our in silico analyses show that basic region can recognize sub sites L and R individually as distinct targets in SH-LR.

**In Vitro Studies via DNase I Footprinting and EMSA.**

Our in silico analyses suggest that the GCN4 basic region can recognize the L and R subsites as distinct targets at SH-LR. We performed in vitro studies to further examine this finding. Therefore, we explored the contribution of each sub site to the affinity between the basic region and SH-LR by analyzing the change in bZIP–DNA affinity when each sub site is eliminated from C/EBP (contains two copies of SH-LR; each SH-LR contains one L and one R sub site): we examined the Kd values of wt bZIP at sub sites L and R, half-site SH-LR, full-site C/EBP, and derivative target sites (Figure 2B and Figure S3 of the Supporting Information). We first investigated DNA-binding activities of wt bZIP at these target sites using DNase I footprinting and EMSA to establish the DNA sequences contacted by each basic region.

We used wt bZIP for two reasons. First, it contains the GCN4 basic region (Figure 1) and mimics the GCN4 bZIP in α-helical structure and DNA-binding function.<sup>7,14,15</sup> We also examined the Kd values of wt bZIP versus GCN4 bZIP at the AP-1 and CRE sites, and at the cognate TGAC half-site and found their DNA-binding functions to be comparable (section S6.4 of the Supporting Information). Second, using wt bZIP allows direct comparison with results reported in our previous work and obtained from the same proteins and techniques,<sup>7,8,21</sup> including e-wt bZIP for footprinting and e- and s-wt bZIPs for EMSA.

The derivative target sites are C/EBP-1, C/EBP-2, XRE1, Arnt E-box, AC, and AC-1. We made a single base-pair T-to-A mutation to the 3′ end of the C/EBP site to generate C/EBP-1, which comprises one SH-LR and one R sequence. We made the same mutation to each end of the C/EBP site to generate C/EBP-2, which comprises two R sequences. XRE1 contains one SH-LR and one TGAC (Arnt E-box half-site), whereas Arnt E-box contains two copies of TCAC. The AC site comprises one TGAC (cognate half-site) and one SH-LR. We made a single base-pair T-to-A mutation to the 3′ end of the AC site to generate AC-1, which contains one TGAC and one R sequence. These T-to-A mutations result in C/EBP-1, C/EBP-2, and AC, which are still flanked by A/T base pairs (Figure 2B); A/T flanking base pairs are preferred by GCN4,23 utilized in the yeast his3 promoter region,<sup>7,2,24</sup> and present in the GCN4–DNA crystal structures.<sup>4–6</sup> All target sites used for Kd analyses are consistently flanked by A/T base pairs; hence, changes in Kd values of bZIP–DNA complexes directly relate to changes in target site sequences.

As shown previously by footprinting and EMSA,<sup>7,8,21</sup> wt bZIP achieves sequence-selective DNA binding at full sites C/EBP, C/EBP-2, XRE1, and Arnt E-box; half-sites SH-LR and Arnt E-box (TCAC); and individual L and R sequences [EMSA conditions were optimized to show sequence-selective DNA binding by wt bZIP (discussed in Section S6.4 of the Supporting Information)]. In this work, EMSA showed that the s-wt bZIP dimer bound to the AC, AC-1, and C/EBP-1 sites in a sequence-selective manner (Figure 4). We assayed other target sites for direct comparison and observed that wt bZIP complexes with some target sites migrated slightly faster than others, potentially because of the basic region’s interaction with SH-LR differing from that at the cognate half-site (discussed in section S2.2 of the Supporting Information). We also observed clear footprints of e-wt bZIP at AC, AC-1, and C/EBP-1 (Figure S4 of the Supporting Information; quantitative phosphorimaging in Figure S5 of the Supporting Information). As with EMSA, the footprinting showed sequence-selective DNA binding of wt bZIP at AC, AC-1, and C/EBP-1.

We determined apparent dimeric Kd values of wt bZIP–DNA complexes by EMSA titrations with s-wt bZIP (Table 1 and Table S7 of the Supporting Information). The net bound
Table 1. Dissociation Constants for the Synthetic wt bZIP in Complex with C/EBP versus Related Target Sites

<table>
<thead>
<tr>
<th>Target Site</th>
<th>$K_d$ (M$^{-1}$)</th>
<th>Contact Locations</th>
<th>Sequence Contacted by</th>
<th>$\Delta K_d^{d}$ (C/EBP)</th>
<th>$\Delta K_d^{d}$ (C/EBP-1)</th>
<th>$\Delta K_d^{d}$ (XRE1)</th>
<th>$\Delta K_d^{d}$ (5H-LR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L sub-site</td>
<td>$&gt;1.0 \times 10^{-11}$</td>
<td>5'-TGGC- 3'</td>
<td>L NA</td>
<td>-</td>
<td>-</td>
<td>&gt;10x</td>
<td></td>
</tr>
<tr>
<td>R sub-site</td>
<td>$&gt;1.0 \times 10^{-11}$</td>
<td>5'-TGGC- 3'</td>
<td>R NA</td>
<td>-</td>
<td>&gt;40x</td>
<td>10x</td>
<td></td>
</tr>
<tr>
<td>5H-LR half-site</td>
<td>$1.1 \pm 0.4 \times 10^{-12}$</td>
<td>5'-TGGC- 3'</td>
<td>L, R (5H-LR) NA</td>
<td>70x</td>
<td>5x</td>
<td>20x</td>
<td>Ref</td>
</tr>
<tr>
<td>C/EBP-2</td>
<td>$2.0 \pm 0.1 \times 10^{-12}$</td>
<td>5'-TGGC- 3'</td>
<td>R, R</td>
<td>8x</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C/EBP-1</td>
<td>$2.4 \pm 0.1 \times 10^{-13}$</td>
<td>5'-TGGC- 3'</td>
<td>L, R (5H-LR) R</td>
<td>16x</td>
<td>Ref</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C/EBP</td>
<td>$1.5 \pm 0.4 \times 10^{-14}$</td>
<td>5'-TGGC- 3'</td>
<td>L, R, L, R (5H-LR)</td>
<td>Ref</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>XRE1</td>
<td>$5.7 \pm 0.9 \times 10^{-14}$</td>
<td>5'-TGGC- 3'</td>
<td>L, R (5H-LR) TCAC</td>
<td>4x</td>
<td>-</td>
<td>Ref</td>
<td>-</td>
</tr>
<tr>
<td>Arnt E-box</td>
<td>$3.2 \pm 0.0 \times 10^{-13}$</td>
<td>5'-TGGC- 3'</td>
<td>TCAC, TCAC</td>
<td>-</td>
<td>-</td>
<td>6x</td>
<td>-</td>
</tr>
<tr>
<td>Arnt E-box half-site</td>
<td>$&gt;1.0 \times 10^{-11}$</td>
<td>5'-TGGC- 3'</td>
<td>TCAC NA</td>
<td>-</td>
<td>&gt;170x</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

“See Figure 2B and Figure S3 of the Supporting Information for DNA duplex sequences. $\Delta K_d$ value is the average of two values from independent EMSA data sets fit to eq S1 with $R$ values of >0.970. See Table S7 of the Supporting Information for $\Delta K_d$ and February 2 of the Supporting Information for representative equilibrium binding isotherms. Sequences contacted by wt bZIP are in bold, highlighted in gray. Sequences recognized by BR$_A$ and BR$_B$ (left and right basic regions in each complex) were confirmed by target site analysis (see section S3 of the Supporting Information). Increase in $K_d$ values as compared with that at the indicated target site (marked as “Ref”). “–”, not compared. From ref 21. From ref 8.

The GCN4 Basic Region Interacts with both Subsites at 5H-LR. C/EBP-2 comprises two R sequences. wt bZIP showed >5-fold stronger affinity at C/EBP-2 than at a single R sequence [$K_d$ values at C/EBP-2 vs sub site R (Table 1)]. This indicates that BR$_A$ and BR$_B$ of the wt bZIP dimer contact one R sequence each at C/EBP-2. Furthermore, C/EBP-1 comprises one 5H-LR and one R sequence. wt bZIP showed a >5-fold stronger affinity at C/EBP-1 than at either 5H-LR or a single R sequence. Therefore, BR$_A$ interacts with 5H-LR at C/EBP-1 when BR$_B$ contacts one R sequence.

wt bZIP at C/EBP-1 or C/EBP-2 differs in that BR$_A$ interacts with 5H-LR (comprising L and R) at C/EBP-1 but contacts only one R sequence at C/EBP-2 (Table 1). wt bZIP also exhibited an 8-fold stronger affinity at C/EBP-1 than at C/EBP-2. This indicates that BR$_A$ contacts not only sub site R at C/EBP-1, as it does at C/EBP-2, but also sub site L. These analyses together show how wt bZIP interacts with C/EBP-1: BR$_A$ interacts with both sub sites at 5H-LR, and BR$_B$ contacts one R sequence.

Arnt E-box comprises two copies of TCAC (Table 1). wt bZIP exhibited an >30-fold stronger affinity at Arnt E-box than at a single TCAC [$K_d$ values at the Arnt E-box vs Arnt E-box half-site (Table 1)]. This indicates that BR$_A$ and BR$_B$ contact TCAC each at Arnt E-box. Moreover, XRE1 contains one 5H-LR and one TCAC. wt bZIP exhibited a >20-fold stronger affinity at XRE1 than at either 5H-LR or TCAC. Hence, BR$_A$ interacts with 5H-LR at XRE1 when BR$_B$ contacts TCAC.

wt bZIP at XRE1 or Arnt E-box differs in that BR$_A$ interacts with 5H-LR (comprising L and R) at XRE1 but contacts TCAC at Arnt E-box (Table 1); L and TCAC are thermodynamically equivalent [they exhibited the same $\Delta K_d$ values, and thus the same affinities, for wt bZIP (Table S7 of the Supporting Information); also, in silico results showed that the GCN4 basic region at L makes base-specific H-bonds only to the end base pairs, which are the same in TCAC]. wt bZIP also exhibited a 6-fold stronger affinity at XRE1 than at Arnt E-box. This indicates that BR$_A$ contacts not only sub site L at XRE1, which is thermodynamically equivalent to contacting TCAC at Arnt E-box, but also sub site R. These analyses together show how wt bZIP interacts with XRE1: BR$_A$ interacts with both sub sites at 5H-LR, and BR$_B$ contacts TCAC.

wt bZIP at C/EBP, C/EBP-1, or XRE1 differs in that BR$_B$ interacts with 5H-LR (comprising L and R) at C/EBP but contacts one R sequence at C/EBP-1, or TCAC (equivalent to L) at XRE1. wt bZIP also exhibited 16- or 4-fold stronger affinity at C/EBP than at C/EBP-1 or XRE1, respectively. This indicates that BR$_B$ interacts with both L and R sub sites at 5H-LR. These analyses together show how wt bZIP interacts with C/EBP: BR$_A$ and BR$_B$ each interact with both sub sites at 5H-LR.

During these $\Delta K_d$ analyses, we found that one sub site in the absence or presence of the other sub site enhances the DNA-
The GCN4 Basic Region Interacts with Either Subsite Individually at 5H-LR. Both in vitro and in silico results suggest that the basic region recognizes either subsite individually at 5H-LR. Our in vitro results also show that the basic region interacts with both subsites at 5H-LR. How does the basic region recognize either subsite but interact with both at 5H-LR? We considered two possibilities. Would the basic region bind to either subsite until it dissociates from 5H-LR, or would the basic region be mobile along 5H-LR and contact the subsites alternately?

We examined the first possibility, which results in a mixture of two populations: some GCN4 basic regions could specifically contact only subsite L, and others specifically contact only subsite R, until they dissociate from 5H-LR. If this is the case, the apparent affinity between the basic region and 5H-LR would lie between the affinities at individual L or R sequences. In fact, wt bZIP exhibited a 10-fold stronger half-site-binding affinity at 5H-LR than at 5H-LR (Table 1); these two sites differ in that BRα contacts one R sequence at C/EBP-1, but nonspecific DNA at 5H-LR. This shows that BRα’s affinity is enhanced because of interaction with this R by 5-fold versus that for nonspecific DNA binding. wt bZIP exhibited 4-fold stronger affinity at C/EBP than at XRE1; these two sites differ in that BRα interacts with both L and R subsites at C/EBP but contacts TCAC (equivalent to L) at XRE1. This shows that BRα’s affinity was enhanced because of interaction with subsite R by 4-fold from binding to only subsite L, similar to the 5-fold increase from nonspecific DNA binding.

Similarly, wt bZIP exhibited a 20-fold stronger affinity at XRE1 than at 5H-LR; BRα contacts TCAC (equivalent to L) at XRE1 but contacts nonspecific DNA at 5H-LR (Table 1). wt bZIP exhibited a 16-fold stronger affinity at C/EBP than at C/EBP-1; BRα interacts with both L and R subsites at C/EBP but contacts R only at C/EBP-1. These comparisons show that BRα’s affinity was enhanced due to interaction with subsite L by 20-fold from nonspecific DNA binding, similar to the 16-fold increase from binding to only subsite R. These analyses show that each subsite enhances the DNA-binding affinity of wt bZIP, in a manner that is independent of the other subsite. This in vitro finding supports the same conclusion from our in silico studies: the basic region recognized either subsite as a distinct target at 5H-LR.

If the basic region is mobile along 5H-LR, the pattern of direct H-bonds made by the basic region with either subsite or with 5H-LR must be changeable. In fact, we found that the pattern differs between snapshots α1em and α2em (GCN4 bZIP complex with C/EBP) and between snapshots β1em and β2em (GCN4 bZIP complex with C/EBP-1): the left halves of α1em and α2em differ by four H-bonds and their right halves by eight H-bonds (Table 2); the left halves of β1em and β2em differ by 12 H-bonds (Table S5 of the Supporting Information) and their right halves by nine H-bonds (Table S6 of the Supporting Information). These results suggest that Asn235 may be mobile along 5H-LR.

### Table 2. Different H-Bonding in α1em versus α2em

<table>
<thead>
<tr>
<th>basic region</th>
<th>DNA</th>
<th>H-bonding</th>
<th>α1em</th>
<th>α2em</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRα</td>
<td>Arg245 N12</td>
<td>A3, OP1</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Arg234 N12</td>
<td>A2, OP2</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Arg245 N12</td>
<td>T18, OP2</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>Arg234 N12</td>
<td>G1, OP1</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Arg234 N12</td>
<td>C16, OP1</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>Arg232 N12</td>
<td>C9, OP1</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

*Atom locants follow IUPAC nomenclature (see section S1 of the Supporting Information for transcription). Residues are numbered according to GCN4 (Figure 1). BRα and BRβ are left and right basic regions in each complex, respectively. See Figure 2C for nucleotide numbering.

H-bonds (Table S5 of the Supporting Information) and their right halves by nine H-bonds (Table S6 of the Supporting Information). These results contradict a single set of contacts, as observed between GCN4 bZIP and DNA in the GCN4–DNA crystal structures.4–8 Moreover, we compared direct H-bonds at the GCN4–DNA interface in crystal structures 1YSA and 1DGC55 versus those in their energy-minimized complexes 1YSAem and 1DGCem, respectively, in control experiments; we found that the H-bonds were maintained after energy minimization in both cases, as shown in section S1.2 of the
Supporting Information. These results suggest that the pattern of direct H-bonds made by the basic region with either subsite or with 5H-LR is variable, which suggests mobility of the basic region along 5H-LR.

We also found distance variations between the same H-bonding pairs at the GCN4–DNA interfaces in ω1,em versus ω2,em and in β1,em versus β2,em (Tables S5 and S6 of the Supporting Information). The distance variations indicate displacements of H-bonding atoms and therefore suggest mobility of these atoms at the GCN4–DNA interfaces. Mobility of these atoms at the interfaces also suggests mobility of the basic region along 5H-LR.

Translocation of the GCN4 Basic Region between Subsites. Our in vitro and in silico results together suggest the mobility of the GCN4 basic region along 5H-LR and thus suggest the possibility for the basic region to recognize the L and R subsites alternately. The basic region must translocate between the subsites to recognize them alternately, given that their positions differ by 1 bp. How is this accomplished?

In fact, DNA-binding proteins translocate along DNA through the genome to search for their target sites. Adam and Delbruck in 1968 formulated a two-stage process: proteins first reach a genomic DNA segment via random diffusion and then translocate along DNA; the second stage reduces dimensionality and thus accelerates target-site search. Riggs et al. in 1970 reported that Lac repressor accomplishes target-site search ~2 orders of magnitude faster than random diffusion. In 1981, Berg et al. developed mathematical descriptions of four mechanisms for rapid protein translocation along DNA: sliding, hopping, jumping, and intersegment transfer. Among these, sliding and hopping are relevant to closely spaced DNA segments and, therefore, relevant to the L and R subsites in 5H-LR.

The sliding mechanism has been studied using various proteins, including transcription factors (for a review, see ref 31). The sliding motion has been captured in vitro by a variety of techniques, including single-molecule AFM and FRET, e.g., sliding of protein Ku was shown by EMSA. The hopping mechanism has also been explored using diverse proteins and captured by various techniques; e.g., hopping of the Hox9 domain was shown by nuclear magnetic resonance (NMR).  

The sliding and hopping mechanisms occur during target-site search. Transcription factors use the same DNA-binding domains to search for and then localize to their target sites. NMR studies of the Hox9 domain indicated that DNA-binding domains employ similar structures to both search for and bind to target sites. Similarly, CD studies presented in our previous work showed similar α-helicity in wt bZIP in the presence of the C/EBP site and a nonspecific DNA sequence. This suggests that the GCN4 basic region can use similar structure, and thus the same protein surface and DNA-binding residues, to search for and bind to target sites. These similarities suggest the possibility for a basic region to alternate between specific binding to its target site and sliding or hopping along DNA.

In fact, many proteins have shown target-site binding followed by sliding along flanking DNA segments, e.g., EcoRI methylase, RNA polymerase, and Lac repressor (for a review, see ref 34). For these proteins, dissociation from target sites has been found to be a two-step process: proteins slide onto DNA segments flanking the target sites and then dissociate into bulk solution, as shown by RNA polymerase. This is contrary to direct dissociation from target sites. This two-step process allows dissociation from target sites to be combined with an association mediated by the sliding mechanism. Nonspecific flanking DNA segments may also act as antennae to collect proteins for later binding to their target sites and may allow proteins to return to their target sites, permitting a transient secondary contact, further stabilizing complexes with target sites. Several studies have found that extending the length of this antenna increases affinity by trapping or attracting the proteins along the DNA duplex for later binding. Surby and Reich found that extending the DNA duplex length from 14 to 775 bp, thus increasing the sliding length, increased target-site affinity by 20-fold for EcoRI methylase.  

Similarly, by extending the DNA duplex length from 36 to 50000 bp, Khoury et al. observed 15-fold affinity increases for Lac repressor. These findings may explain how the sliding mechanism increases affinities between proteins and their target sites, and why we observed affinity increases when the GCN4 basic region interacts with 5H-LR.

We suggest this explanation because of the following. If the basic region only binds to and directly dissociates from individual L and R subsites, the affinity of the basic region at 5H-LR should lie between the affinities at either the L or R sequence. However, wt bZIP exhibited 10-fold stronger half-site binding affinity at 5H-LR than at either the L or R sequence, as shown previously (Table 1); similar results were obtained when we analyzed above the Kd values at C/EBP versus 5H-LR, C/EBP versus C/EBP-1, and C/EBP versus XRE1. These results suggest that other factors are involved in further increasing the affinity of this interaction. We note that in the cases of EcoRI methylase and Lac repressor, proteins sliding onto nonspecific DNA, even with weak affinity at individual nonspecific DNA segments, exhibit much higher affinity at their target sites. However, in our case, the basic region translocates between subsites where the basic region already exhibits an affinity higher than that of nonspecific DNA binding. The further enhanced affinity suggests that the basic region can slide between subsites, which increases overall affinity at 5H-LR.

Would the basic region also hop between subsites whose positions differ by 1 bp? Wunderlich and Mirny have discussed that for some proteins, the theoretical prediction indicates a median hopping distance of ~1 bp; such a short distance is within the observational limitations of various experimental techniques and thus would be missed by single-molecule experiments. For this reason, the authors suggest that within such a short distance, hopping could be considered equivalent to sliding. Furthermore, Winter et al. estimated a 100 bp sliding length for Lac repressor before dissociation from DNA. In SH-LR, the subsite positions differ by only 1 bp. This suggests that the basic region is likely to slide between subsites at SH-LR.

A Highly Dynamic DNA-Binding Model for the GCN4 Basic Region To Interact with 5H-LR. In this work, we explored how the GCN4 basic region interacts with the SH-LR half-site. We investigated the interface between the basic region and SH-LR by analyzing snapshots of the interface generated by AMBER simulation. We analyzed Kd values of wt bZIP complexes with target sites that contain SH-LR versus either subsite; we analyzed Kd differences to investigate how the L and R subsites contribute to affinity between the basic region and SH-LR. The in vitro and in silico results offer the following insights into how the basic region interacts with SH-LR.
Our results suggest that the basic region does not recognize SH-LR solely as a single target site but that it can recognize subsites individually as distinct targets at SH-LR. The basic region may dissociate from one subsite and reassociate with the other at SH-LR and may translocate between the subsites, potentially by sliding and hopping. These results together suggest a highly dynamic DNA-binding model for the basic region to interact with SH-LR.

Our results also show that when one basic region translocates along SH-LR, the partner basic region can engage in various DNA-binding activities: (i) at C/EBP, the partner basic region also interacts with SH-LR in the same way; (ii) at C/EBP-1 and XRE1, the partner basic region engages in weak but sequence-selective DNA binding at the noncognate half-site; (iii) at AC, the partner basic region engages in strong and sequence-specific DNA binding at the cognate half-site; and (iv) at a single SH-LR, the partner basic region engages in nonspecific DNA binding. Therefore, the two basic regions of the bZIP dimer may behave as monomers because they can independently engage in different DNA-binding activities, including sliding along SH-LR. Several in vitro studies have already supported the monomer pathway for dimeric proteins, including GCN4, to form complexes with DNA. In this pathway, monomers associate with target sites independently, as bZIP basic regions do, before dimerization at the target site. Our findings are consistent with the monomer pathway in explaining not only how basic regions associate with DNA but also how they slide along DNA in vivo.

At C/EBP, C/EBP-1, XRE1, and AC, a basic region translocates along SH-LR as its partner engages in various DNA-binding activities. This will require flexibility in bZIP α-helical structure. To compare backbone motion in the DNA-bound versus free GCN4 bZIP, Columbus and Hubbell placed nitroxide spin-labels on the solvent-exposed surface of the GCN4 bZIP α-helix and performed solution EPR studies. Their EPR studies demonstrated that even when bound to cognate AP-1, the GCN4 basic region exhibited significant mobility in its backbone, although motion is reduced, compared to free GCN4 bZIP. The EPR studies also suggested that axial twisting originating from the hinge between the basic region and leucine zipper could constitute a rigid-body axial rocking motion of the basic region. Such rocking may also permit a rocking motion of the basic region. This will require backbone and positively charged protein side chains. As discussed above, DNA-DNA interactions switch to sequence-specific interactions, including base-specific H-bonds for DNA interactions.

In addition, the EPR results are consistent with solution NMR studies performed on GCN4 bZIP, which demonstrate that the basic region is highly dynamic. Interestingly, only free GCN4 bZIP has been characterized by NMR, whereas high-resolution information about the bZIP–DNA complex has not been achieved by solution methods, but by X-ray crystallography. In their NMR studies, Palmer and co-workers observed that changes in conformational dynamics of the basic region backbone occur upon binding to DNA and contribute favorably to the overall thermodynamics of complex formation. Likewise, we found that basic region translocation promotes binding affinity in the bZIP–DNA complex.

Aguado-Llera et al. found the structure of the basic helix-loop-helix (bHLH) domain of human neurogenin 1 bound to the E-box to be "fuzzy": the protein–DNA complex displayed...
sites by first sliding onto flanking DNA segments. These observations point to the possibility that these proteins can slide on and off their target sites in vivo. Similarly, the basic region can slide on and off either subsite in our dynamic DNA-binding model. Our analysis described above regarding wt bZIP at C/EBP, C/EBP-1, XRE1, and AC sites suggests that for dimeric DNA-binding proteins, one DNA-binding domain can slide on and off its target site while the partner DNA-binding domain engages in various DNA-binding activities. We speculate that some transcription factors may use this dynamic DNA-binding model during their target-site search and localization tasks in vivo.

Conclusion. This work adds further understanding of how bZIP transcription factors interact with their cognate gene regulatory sequences in vivo. We accomplished this goal through in vitro and in silico studies on noncognate but sequence-selective DNA binding by the bZIP domain of transcription factor GCN4. Our results show that the bZIP basic region may not always recognize a half-site as a single target: a half-site may comprise shorter subsites. The basic region can recognize individually. Our results suggest a highly dynamic DNA-binding model for bZIP transcription factors to interact with their target sites: in a case where a half-site comprises a subsite, the basic region alternately recognizes the subsites as distinct targets by translocating between them via sliding and hopping. The basic region is mobile if this model is used; however, the bZIP transcription factor is still localized to the cognate gene regulatory sequences with high specificity and affinity. This model may also be useful during the transition from genomic target-site search to interaction between bZIP transcription factors and their cognate gene regulatory sequences. Although translocation between subsites was not directly observed and may be further investigated by NMR or EPR, this work provides evidence to support this DNA-binding model and adds additional understanding of how bZIP transcription factors search for and localize to their cognate gene regulatory sequences.

■ ASSOCIATED CONTENT

Supporting Information

Additional details for molecular modeling, qualitative references of $K_d$ values, target-site analyses, analysis of the wt bZIP complex with the AC site, and summary of protocols of in vitro studies. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

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

Funding

This work was supported by 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.

Notes

The authors declare no competing financial interests.

■ ACKNOWLEDGMENTS

We thank Ulrich Krull for providing access and funding for molecular modeling; Lakshmi Kotra; William Wei, and the AMBER community, especially Ross Walker and Mark Williamson, for advice on modeling; Alevtina Pavlenco for technical assistance; and Cherie Werhun for helpful discussions about writing the manuscript.

■ ABBREVIATIONS

SH-LR, 6 bp hybrid of subsites L and R; AC, hybrid of the AP-1 and C/EBP sites; AFM, atomic force microscopy; AMBER, Assisted Model Building with Energy Refinement; Arnt, aryl hydrocarbon receptor nuclear translocator; bHLH, basic helix-loop-helix; BR$_{	ext{o}}$, basic regions in the left half of the bZIP–DNA complex; BR$_{	ext{p}}$, basic regions in the right half of the bZIP–DNA complex; BSA, bovine serum albumin; bZIP, basic region/leucine zipper; CAII, bovine carbonic anhydrase II; CD, circular dichroism; C/EBP, CCAAT/enhancer binding protein; CMDP, Center for Molecular Design and Preformulations; Coot, Crystallographic Object-Orinted Toolkit; CRE, cAMP response element; DTT, dithiothreitol; E-box, enhancer box; EDTA, ethylenediaminetetraacetic acid; EMSA, electrophoretic mobility shift assay; EPR, electron paramagnetic resonance; ESI-MS, electrospray ionization mass spectrometry; FRET, fluorescence resonance energy transfer; HPLC, high-performance liquid chromatography; IIF, integration host factor; IUPAC, International Union of Pure and Applied Chemistry; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; T$\alpha$-leap, temperature leap; UV–vis, ultraviolet–visible; VMD, Visual Molecular Dynamics; wt bZIP, wild-type bZIP; e-wt bZIP, bacterially expressed wt bZIP; s-wt bZIP, chemically synthesized wt bZIP; XRE1, xenobiotic response element 1; $K_d$, apparent dimeric dissociation constant; $\Delta\theta_{\text{app}}$, net bound DNA fraction.

■ REFERENCES