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 Shahravan,[†] Anna V. Fedorova,[†] and Jumi A. Shin*[,†,‡]
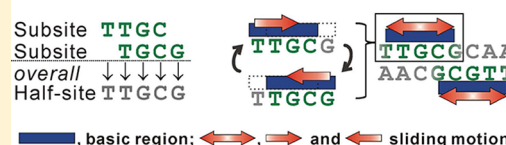
[†]Department of Chemistry, University of Toronto, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada

[‡]Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College Street, Toronto, Ontario M5S 3G9, Canada

S *Supporting Information*

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 5H-LR) and that 5H-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 5H-LR when the bZIP dimer localizes to full-sites. Using AMBER molecular modeling, we simulated GCN4 bZIP complexes with full-sites containing 5H-LR to investigate in silico the interface between the basic region and 5H-LR. We also performed in vitro investigation of bZIP–DNA interactions at a number of full-sites that contain 5H-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 5H-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.

Transcription 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 dimer targets cognate full-site AP-1 (7 bp, TGACTCA), 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.[5,6] Crystal structures of this bZIP with AP-1 or CRE show that each basic region makes a static set of base-specific contacts with 4 bp TGA(C/G),[4–6] indicating that the basic region recognizes the 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. We studied sequence-selective bZIP–DNA interactions to explore how bZIP proteins search for and localize to their target sites. Consequently, we found that 4 bp TTGC and TGCG are noncognate target sites of the GCN4 basic region and can overlap to give 5 bp TTGCG.[7,8] Therefore, we named 4 bp TTGC as "L" (left) and TGCG as "R" (right), the 5′ and 3′ subsites of 5-bp TTGCG, respectively; we termed this 5 bp sequence "5H-LR" as a 5 bp hybrid of subsites L and R. Additionally, the basic region exhibited ≥10-fold higher affinity at 5H-LR than that at either L

(A) The GCN4 bZIP   GCN4 basic region   GCN4 leucine zipper
         D P A A L K R A R N T E A A R R S R A R K L Q R M K Q – <u>L E D K V E E L L S K N Y H L E N E V A R L K K L V G E R</u>

(B) The wt bZIP   D P A A L K R A R N T E A A R R S R A R K L Q R M K Q – <u>L E Q K V L E L T S D N D R L R K R V E Q L S R E L D T L</u>
            GCN4 basic region   C/EBP leucine zipper

Expressed wt bZIP (e-wt bZIP, 96-mer)   *M G G S H H H H H H G M A S M T G G Q Q M G R D L Y D D D D K –*
                                   GCN4 basic region-<u>C/EBP leucine zipper</u>– *G G C G G Y Y Y Y*

Synthetic wt bZIP (s-wt bZIP, 57-mer)   **GCN4 basic region**-<u>C/EBP leucine zipper</u>– *Y*

Figure 1. Protein sequences. Basic regions are shown in bold with leucine zippers underlined. (A) Native GCN4 bZIP (residues 226–281).[49] (B) wt bZIP: the GCN4 basic region (residues 226–252) fused to C/EBP leucine zipper (residues 310–338).[14] The N-terminal Met of e-wt bZIP was removed during post-translational modification.[50]

(A) Sequences used in footprinting analysis

AP-1   5′–TCCCTCCGGA<u>**TGACTCA**</u>TTTTTTGCGA–3′
C/EBP-1   5′–CCTGCAGGAA<u>**TTGCGCA**</u>TTGAAGGTTC–3′
AC   5′–CCTGCAGGAA<u>**TGACGCAA**</u>TGAAGGTTC–3′
AC-1   5′–CCTGCAGGAA<u>**TGACGCA**</u>TTGAAGGTTC–3′

(B) Sequences used in EMSA

AP-1   5′–TGCAGGAA<u>**TGACTCA**</u>TTGAAGGTT–3′
CRE   5′–TGCAGGAA<u>**TGACGTCA**</u>TGAAGGTT–3′
XRE1   5′–TGCAGGAA<u>**TTGCGTGA**</u>TGAAGGTT–3′
AC   5′–TGCAGGAA<u>**TGACGCAA**</u>TGAAGGTT–3′
AC-1   5′–TGCAGGAA<u>**TGACGCA**</u>TTGAAGGTT–3′
C/EBP   5′–TGCAGGAA<u>**TTGCGCAA**</u>TGAAGGTT–3′
C/EBP-1   5′–TGCAGGAA<u>**TTGCGCA**</u>TTGAAGGTT–3′
C/EBP-2   5′–TGCAGGAA<u>**ATGCGCA**</u>TTGAAGGTT–3′
Arnt E-box 5′–TGCAGGAA<u>**TCACGTGA**</u>TGAAGGTT–3′
AP-1H24   5′–TGCAGGAA<u>**ATGACTTT**</u>TGAAGGTT–3′
5H-LR   5′–TGCAGGAA<u>**ATTGCG**</u>TTTGAAGGTT–3′

(C) Sequences used in molecular modeling

AP-1   1 2 3 4 5 6 7 8 9 10 11 12 13 14
5′- T T C C T A **T G A C T C A** T C C A G T T
3′- A G G A T **A C T G A G T** A G G T C A A A
          28 27 26 25 24 23 22 21 20 19 18 17 16 15

CRE   1 2 3 4 5 6 7 8 9 10 11 12 13 14
5′- T G G A G A **T G A C G T C A** T C T C C
3′- C C T C T **A C T G C A G T** A G A G G T
          28 27 26 25 24 23 22 21 20 19 18 17 16 15

C/EBP   1 2 3 4 5 6 7 8 9 10 11 12 13 14
5′- C A G G A A **T T G C G C A A** T G A A G G
3′- T C C T T **A A C G C G T T** A C T T C C A
          28 27 26 25 24 23 22 21 20 19 18 17 16 15

C/EBP-1   1 2 3 4 5 6 7 8 9 10 11 12 13 14
5′- C A G G A A **T T G C G C A** T T G A A G G
3′- T C C T T **A A C G C G T** A A C T T C C A
          28 27 26 25 24 23 22 21 20 19 18 17 16 15

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;[24] 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.[21] For the same reason, core target sites in the AP-1H24 and 5H-LR duplexes were shifted by 1 bp toward the 3′ end.[8] Other duplexes are listed in Figure S3 of the Supporting Information.[7,8] 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.[4,5] C/EBP and C/EBP-1 are same as those for the EMSA.

---

or R.[8] 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 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;[9] this suggests the possibility that the basic region can be mobile along 5H-LR. DNA-binding proteins also exhibit mobility by translocating along genomic DNA during target-site search.[10] There are four mechanisms for rapid protein translocation along DNA:[11] (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)[12] and hopping of human DNA repair factor RAD54 by atomic force microscopy (AFM).[13] 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;[7,14,15] 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 5H-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
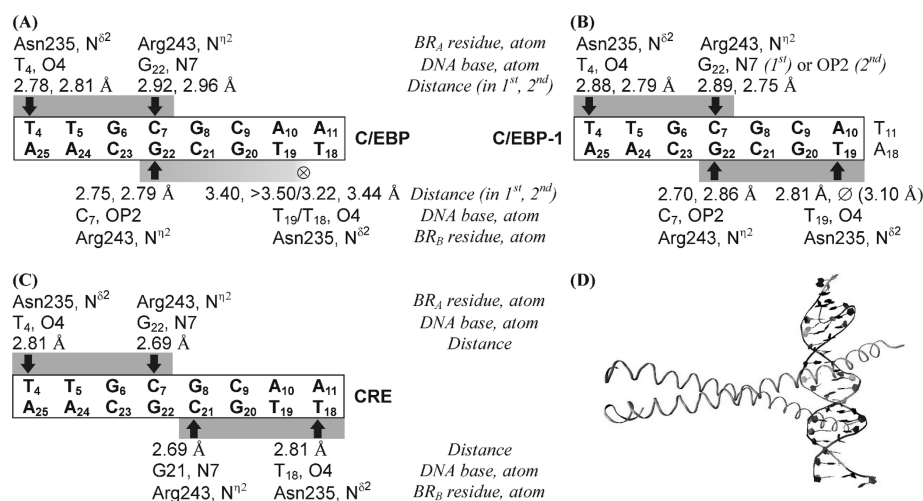
(A)

Asn235, N$^{\delta 2}$	Arg243, N$^{\eta 2}$		*BR$_A$ residue, atom*
T$_4$, O4	G$_{22}$, N7		*DNA base, atom*
2.78, 2.81 Å	2.92, 2.96 Å		*Distance (in 1st, 2nd)*

T$_4$	T$_5$	G$_6$	C$_7$	G$_8$	C$_9$	A$_{10}$	A$_{11}$	C/EBP
A$_{25}$	A$_{24}$	C$_{23}$	G$_{22}$	C$_{21}$	G$_{20}$	T$_{19}$	T$_{18}$	

2.75, 2.79 Å   3.40, >3.50/3.22, 3.44 Å   *Distance (in 1st, 2nd)*
C$_7$, OP2   T$_{19}$/T$_{18}$, O4   *DNA base, atom*
Arg243, N$^{\eta 2}$ Asn235, N$^{\delta 2}$ *BR$_B$ residue, atom*

(B)

Asn235, N$^{\delta 2}$	Arg243, N$^{\eta 2}$	
T$_4$, O4	G$_{22}$, N7 *(1st)* or OP2 *(2nd)*	
2.88, 2.79 Å	2.89, 2.75 Å	

T$_4$	T$_5$	G$_6$	C$_7$	G$_8$	C$_9$	A$_{10}$	T$_{11}$	C/EBP-1
A$_{25}$	A$_{24}$	C$_{23}$	G$_{22}$	C$_{21}$	G$_{20}$	T$_{19}$	A$_{18}$	

2.70, 2.86 Å   2.81 Å, ∅ (3.10 Å)   *Distance (in 1st, 2nd)*
C$_7$, OP2   T$_{19}$, O4   *DNA base, atom*
Arg243, N$^{\eta 2}$ Asn235, N$^{\delta 2}$ *BR$_B$ residue, atom*

(C)

Asn235, N$^{\delta 2}$	Arg243, N$^{\eta 2}$		*BR$_A$ residue, atom*
T$_4$, O4	G$_{22}$, N7		*DNA base, atom*
2.81 Å	2.69 Å		*Distance*

T$_4$	T$_5$	G$_6$	C$_7$	G$_8$	C$_9$	A$_{10}$	A$_{11}$	CRE
A$_{25}$	A$_{24}$	C$_{23}$	G$_{22}$	C$_{21}$	G$_{20}$	T$_{19}$	T$_{18}$	

2.69 Å   2.81 Å   *Distance*
G21, N7   T$_{18}$, O4   *DNA base, atom*
Arg243, N$^{\eta 2}$ Asn235, N$^{\delta 2}$ *BR$_B$ residue, atom*

(D)

Figure 3. Direct H-bonds indicate DNA sequences recognized by the basic regions in (A) $\alpha_{1\text{-em}}$ and $\alpha_{2\text{-em}}$, different snapshots of the GCN4 complex with the C/EBP site, (B) $\beta_{1\text{-em}}$ and $\beta_{2\text{-em}}$, different snapshots of the GCN4 complex with the C/EBP-1 site, and (C) GCN4−CRE crystal structure 1DGC.[5] (D) Snapshot $\alpha_{1\text{-em}}$. These H-bonds are established by N$^{\delta 2}$ of Asn235 and N$^{\eta 2}$ of Arg243. The locations of H-bonds are denoted with arrows. ⊗ indicates the location of N$^{\delta 2}$ of Asn235, which establishes no H-bond. ∅ indicates that N$^{\delta 2}$ of Asn235 is near O4 of T$_{19}$ but is slightly too distant to establish a H-bond.

target of the C/EBP bZIP; this site comprises two copies of 5H-LR), half-site 5H-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 5H-LR and the possibility of basic region sliding and hopping. Our results provide added insights into how α-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 ff99SB 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 α-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 neutrality.[17] 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$^{\delta 2}$ = 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 for an examination of control experiments.

To create α_1 and β_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 α_1 and β_1 were simulated to yield energy-minimized $\alpha_{1\text{-em}}$ and $\beta_{1\text{-em}}$, 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 $\alpha_{1\text{-em}}$ and $\beta_{1\text{-em}}$, we pursued a "reverse" approach to obtain additional snapshots of the same complexes. We utilized $\alpha_{1\text{-em}}$ and restrained four distances using the "makeDIST_RST" function in AMBER: distances from N$^{\delta 2}$ of Asn235 in BR$_A$ and BR$_B$ (basic regions in the left and right halves, respectively, of the bZIP−DNA complex) to O4 of the T$_4$ and T$_{18}$ bases, respectively, were restrained to 2.30−2.80 Å, while distances from N$^{\delta 2}$ of Asn235 in BR$_A$ and BR$_B$ to O4 of the T$_5$ and T$_{19}$ 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$^{\delta 2}$ of Asn235 in BR$_A$ to O4 of the T$_4$ and T$_5$ bases in the left half of $\alpha_{1\text{-em}}$. This simulation was run for 20000 steps, without periodic boundary conditions and with a distance-dependent dielectric (option eedmeth = 5), to create initial structure α_2 (with C/EBP). These distance restraints made α_2 different from α_1 (see the analysis and comparison of initial structures α_1 and α_2 in section S1.1 of the Supporting Information). To create initial structure β_2 (with C/EBP-1), we modified DNA bases 11 and 18 of α_2 (Figure 2C); α_2 and β_2 were otherwise identical. We simulated α_2 and β_2 to obtain energy-minimized $\alpha_{2\text{-em}}$ and $\beta_{2\text{-em}}$ (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$_{em}$, $\alpha_{1\text{-em}}$, and $\beta_{1\text{-em}}$. Unlike the α_2 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 5H-LR half-site, we first examined whether the basic region recognizes 5H-LR solely as a single DNA target or recognizes either subsite L or R as part of its interaction with 5H-LR. We therefore explored the interface between the basic region and 5H-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 5H-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 5H-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$_{em}$ and 1DGC$_{em}$, respectively, for control experiments (see section S1.2 of the Supporting Information for an examination of control experiments). We generated energy-minimized α_{1-em} and α_{2-em}, which are two different snapshots of the complex with C/EBP, as well as β_{1-em} and β_{2-em}, which are two different snapshots of the complex with C/EBP-1 (Figure 3 and Figure S1 of the Supporting Information; different snapshots were generated via different approaches), to explore the interface between the basic region and 5H-LR.

In snapshots α_{1-em}, α_{2-em}, β_{1-em}, and β_{2-em}, the GCN4 bZIP appeared as a dimer of continuous α-helices (Figure 3 and Figure S1 of the Supporting Information), as it does in crystal structure 1YSA (with AP-1). The similar bZIP α-helical structures shown by snapshots α_{1-em} and α_{2-em} (both with C/EBP) and crystal structure 1YSA are consistent with the finding that wt bZIP exhibited similar helicities at C/EBP and AP-1 (65 and 74%, respectively), as shown by the circular dichroism (CD) study presented previously.[7] This indicates that the GCN4 basic region uses a similar conformation to interact with 5H-LR (in C/EBP) and the cognate half-site (in AP-1). Therefore, we further explored the contacts, specifically direct hydrogen bonds (H-bonds), between the basic region and 5H-LR in the snapshots with C/EBP or C/EBP-1.

We explored direct H-bonds for two reasons. First, in the GCN4−DNA crystal structures, 9 of the 12 or 13 DNA-binding residues of the basic regions make direct H-bonds with DNA.[4−6] Hence, direct H-bonds at the GCN4−DNA interface collectively provide a significant representation of the interactions between the basic regions and DNA in each snapshot. Second, direct H-bonds made by N$^{\delta 2}$ of Asn235 and by N$^{\eta 2}$ of Arg243 with DNA indicate the 5′ and 3′ ends, respectively, of the DNA sequences recognized by the basic region. View the left half of crystal structure 1DGC as an

example: N$^{\delta 2}$ of Asn235 and N$^{\eta 2}$ of Arg243 each donate a H-bond to the 5′ and 3′ ends of 4 bp TGAC, respectively (Figure 3); BR$_A$ (basic region in the left half of the complex) makes no direct H-bond that is base-specific beyond these two ends, indicating that BR$_A$ recognizes 4 bp TGAC. Likewise, BR$_B$ (basic region in the right half of the complex) recognizes 4 bp TGAC in the left half of the same crystal structure. Similar analyses also show that each basic region recognizes a cognate half-site, 4 bp TGA(C/G), in crystal structures 1YSA[4] and 2DGC[6] (also the GCN4 bZIP with CRE).

We omitted other intermolecular interactions from our analyses for the following reasons. Including water-mediated H-bonds made by Lys231 and Asn235 expands the recognized DNA sequence from 4 bp TGA(C/G) to 5 bp ATGA(C/G). However, Lys231 and Asn235 contact this additional base pair inconsistently: Lys231 from BR$_A$ does not contact this base pair in 1YSA,[4] whereas BR$_A$ and BR$_B$ use Lys231 to contact the DNA backbone at this base pair in 1DGC[5] but use Asn235 to make base-specific contact with this base pair in 2DGC.[6] Moreover, although this base pair is preferred by the GCN4 bZIP, it is not highly conserved and therefore not considered part of the core target sequence.[22] Hence, we did not include water-mediated H-bonds in the analyses to find 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 α_{1-em}, α_{2-em}, β_{1-em}, and β_{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 α_{1-em}, α_{2-em}, β_{1-em}, and β_{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 5H-LR and the cognate half-sites.

In the following section, we analyzed the DNA sequence recognized by each basic region in snapshots α_{1-em}, α_{2-em}, β_{1-em}, and β_{2-em}. We used direct H-bonds donated by N$^{\delta 2}$ of Asn235 and N$^{\eta 2}$ 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 5H-LR solely as a single DNA target or recognizes either subsite as part of its interaction with 5H-LR.

DNA Sequences Recognized in Snapshots of the GCN4 bZIP Complex with the C/EBP or C/EBP-1 Site. We examined snapshots α_{1-em} and α_{2-em} of the complex with C/EBP, whose left and right halves each contained 5H-LR. These

snapshots differed in their initial structures: BR_B recognized 4 bp subsite R in α_1 versus 5 bp 5H-LR in α_2 (see analysis of α_1 vs α_2 in section S1.1 of the Supporting Information). In the left halves of both snapshots, BR_A recognized 4 bp subsite L but not 5 bp 5H-LR, as indicated by H-bonds made by $N^{\delta2}$ of Asn235 and $N^{\eta2}$ of Arg243 (Figure 3). In the right halves of both snapshots, $N^{\eta2}$ of Arg243 from BR_B contacted the 3′ end of 5H-LR when $N^{\delta2}$ of Asn235 was situated between the two 5′ end base pairs, but $N^{\delta2}$ of Asn235 did not reach the 5′ end of 5H-LR. We found that no basic region recognized the full length of 5H-LR in snapshots $\alpha_{1\text{-em}}$ and $\alpha_{2\text{-em}}$.

We also examined snapshots $\beta_{1\text{-em}}$ and $\beta_{2\text{-em}}$ of the complex with C/EBP-1. The left half of each snapshot contained 5H-LR, but BR_A recognized only 4 bp subsite L (Figure 3). Thus, we found that BR_A did not recognize the full length of 5H-LR in snapshots $\beta_{1\text{-em}}$ and $\beta_{2\text{-em}}$.

These results from snapshots $\alpha_{1\text{-em}}$, $\alpha_{2\text{-em}}$, $\beta_{1\text{-em}}$, and $\beta_{2\text{-em}}$ suggest that the basic region can recognize 4 bp subsite L as a distinct target at 5H-LR. Moreover, the right half of the complex with C/EBP-1 contained one R sequence. BR_B recognized this R sequence in snapshot $\beta_{1\text{-em}}$ (Figure 3); also, $N^{\delta2}$ of Asn235 and $N^{\eta2}$ of Arg243 from BR_B in snapshots $\beta_{1\text{-em}}$ and $\beta_{2\text{-em}}$ exhibited a very similar H-bonding pattern. Beyond this R sequence, BR_B 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 subsite R as a distinct target, regardless of neighboring base pairs, at 5H-LR. These analyses together suggest that the basic region does not recognize 5H-LR solely as a single target; these analyses offer no evidence of recognition of the full length of 5H-LR (i.e., recognition of both subsites simultaneously). However, our in silico analyses show that basic region can recognize subsites L and R individually as distinct targets in 5H-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 5H-LR. We performed in vitro studies to further examine this finding. Therefore, we explored the contribution of each subsite to the affinity between the basic region and 5H-LR by analyzing the change in bZIP−DNA affinity when each subsite is eliminated from C/EBP (contains two copies of 5H-LR; each 5H-LR contains one L and one R subsite): we examined the K_d values of wt bZIP at subsites L and R, half-site 5H-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.[7,14,15] We also examined the K_d 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,[7,8,21] 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 5H-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 5H-LR and one TCAC (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 5H-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,[22,24] and present in the GCN4−DNA crystal structures.[4−6] All target sites used for K_d analyses are consistently flanked by A/T base pairs; hence, changes in K_d values of bZIP−DNA complexes directly relate to changes in target site sequences.

As shown previously by footprinting and EMSA,[7,8,21] wt bZIP achieves sequence-selective DNA binding at full-sites C/EBP, C/EBP-2, XRE1, and Arnt E-box; half-sites 5H-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
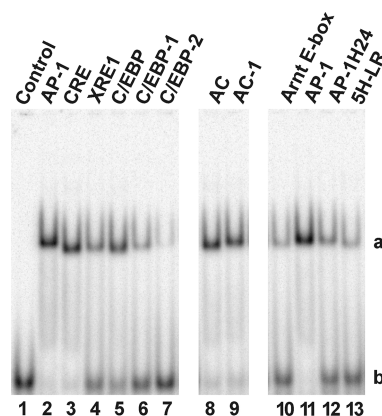

Figure 4. Qualitative EMSA analysis. This figure is from a single gel with additional lanes removed: lane 1, free AP-1 DNA control; lanes 2−13, DNA duplexes with 1 μM s-wt bZIP. AP-1 serves as the specific control for full-site DNA binding and AP-1H24 for half-site DNA binding at 5H-LR.[7,8,21] Note that AP-1H24 (24 bp,[8] Figure 2B) differs from the AP-1 half-site duplex (20 bp,[7] Figure S3 and Table S17 of the Supporting Information), used for K_d determination. Bands corresponding to dimeric wt bZIP complexes with DNA are indicated with "a" and those of free DNA with "b".

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 5H-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 K_d 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

Target Site [a]	K_d (M^2) [b]	Contact Locations [c]	Sequence Contacted by BR$_A$ [c]	BR$_B$ [c]	ΔK_d [d] (C/EBP)	ΔK_d [d] (C/EBP-1)	ΔK_d [d] (XRE1)	ΔK_d [d] (5H-LR)
L subsite	$>1.0 \times 10^{-11}$ [f,g]	5′-**TTGC**---- / 3′-**AACG**----	L	NA	-	-	-	>10x
R subsite	$\gg 1.0 \times 10^{-11}$ [g]	5′--**TGCG**--- / 3′--**ACGC**---	R	NA	-	>40x	-	10x
5H-LR half-site	$1.1 \pm 0.4 \times 10^{-12}$ [g]	5′-**TTGCG**--- / 3′-**AACGC**---	L, R (5H-LR)	NA	70x	5x	20x	**Ref**
C/EBP-2	$2.0 \pm 0.1 \times 10^{-12}$ [g]	5′--**TGCG**CA- / 3′--AC**GCGT**-	R	R	-	8x	-	-
C/EBP-1	$2.4 \pm 0.1 \times 10^{-13}$	5′-**TTGCG**CA- / 3′-**AAC**GCGT-	L, R (5H-LR)	R	16x	**Ref**	-	-
C/EBP	$1.5 \pm 0.4 \times 10^{-14}$ [e,g]	5′-**TTGCG**CAA / 3′-**AAC**GCGTT	L, R (5H-LR)	L, R (5H-LR)	**Ref**	-	-	-
XRE1	$5.7 \pm 0.9 \times 10^{-14}$ [e,f]	5′-**TTGCG**TGA / 3′-**AACG**CACT	L, R (5H-LR)	TCAC	4x	-	**Ref**	-
Arnt E-box	$3.2 \pm 0.0 \times 10^{-13}$ [e,g]	5′-**TCAC**GTGA / 3′-AGTG**CACT**	TCAC	TCAC	-	-	6x	-
Arnt E-box half-site	$>1.0 \times 10^{-11}$ [f,g]	5′-**TCAC**---- / 3′-**AGTG**----	TCAC	NA	-	-	>170x	-

[a]See Figure 2B and Figure S3 of the Supporting Information for DNA duplex sequences. [b]Each 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 $\Delta\theta_{app}$ values and Figure S2 of the Supporting Information for representative equilibrium binding isotherms. [c]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). [d]Increase in K_d values as compared with that at the indicated target site (marked as "Ref"); "−", not compared. [e]From ref 21. [f]From ref 7 [g]From ref 8

DNA fractions, $\Delta\theta_{app}$ (Table S7 of the Supporting Information), and representative binding isotherms (Figure S2 of the Supporting Information) are given as qualitative references of the K_d values (see explanation of $\Delta\theta_{app}$ in section S5.4 of the Supporting Information).

By analyzing the results from footprinting and EMSA (section S3 of the Supporting Information), we confirmed (i) no adventitious sequences targeted by the basic region in any DNA duplex used for K_d determination and (ii) the DNA sequences contacted by each basic region within C/EBP, C/EBP-1, C/EBP-2, XRE1, Arnt E-box, AC, and AC-1. The contacted DNA sequences within these target sites are listed in Table 1, except those within AC and AC-1 are listed in Table S17 of the Supporting Information. In the following sections, we analyzed the data presented in Table 1 to explore how the basic region contacts either subsite and interacts with 5 bp TTGCG at 5H-LR.

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 subsite R (Table 1)].[8] 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 subsite R at C/EBP-1, as it does at C/EBP-2, but also subsite L. These analyses together show how wt bZIP interacts with C/EBP-1:

BR$_A$ interacts with both subsites 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)].[8] 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 $\Delta\theta_{app}$ 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 subsite L at XRE1, which is thermodynamically equivalent to contacting TCAC at Arnt E-box, but also subsite R. These analyses together show how wt bZIP interacts with XRE1: BR$_A$ interacts with both subsites 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 subsites at 5H-LR. These analyses together show how wt bZIP interacts with C/EBP: BR$_A$ and BR$_B$ each interact with both subsites at 5H-LR.

During these K_d analyses, we found that one subsite in the absence or presence of the other subsite enhances the DNA-

binding affinity of wt bZIP to the same degree, as shown in the following. wt bZIP exhibited a 5-fold stronger affinity at C/EBP-1 than at 5H-LR (Table 1); these two sites differ in that BR_B contacts one R sequence at C/EBP-1, but nonspecific DNA at 5H-LR. This shows that BR_B'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 BR_B interacts with both L and R subsites at C/EBP but contacts TCAC (equivalent to L) at XRE1. This shows that BR_B'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; BR_B 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; BR_B interacts with both L and R subsites at C/EBP but contacts R only at C/EBP-1. These comparisons show that BR_B'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.

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 either the L or the R sequence (Table 1), as shown in our previous work.[8] Also, C/EBP and C/EBP-1 differ by a subsite L; C/EBP contains this L subsite, which increases wt bZIP's affinity by 16-fold. C/EBP and XRE1 differ by a subsite R essentially (L and TCAC are thermodynamically equivalent); C/EBP contains this R subsite, which enhances wt bZIP's affinity by 4-fold. By comparison, C/EBP and 5H-LR differ by one additional copy of 5H-LR; C/EBP contains this 5H-LR, which increases wt bZIP's affinity by 70-fold.

These results demonstrate that the affinity between the basic region and 5H-LR is much higher than that at either subsite. These results do not negate the possibility that the basic region can dissociate from one subsite and reassociate with the other at 5H-LR. However, these results show that the basic region binding to either subsite until dissociation from 5H-LR does not sufficiently explain how the basic region interacts with 5H-LR.

If the basic region is mobile along 5H-LR, $N^{\delta2}$ of Asn235, which contacts the 5′ end of DNA sequences recognized by the basic region, not only would contact the 5′ ends of subsite L or R of 5H-LR but also would appear between subsite L's and R's 5′ ends. We observed this from snapshots $\alpha_{1\text{-em}}$, $\alpha_{2\text{-em}}$, $\beta_{1\text{-em}}$, and $\beta_{2\text{-em}}$: $N^{\delta2}$ of Asn235 (i) contacted the 5′ end of subsite L in the left halves of snapshots $\alpha_{1\text{-em}}$, $\alpha_{2\text{-em}}$, $\beta_{1\text{-em}}$, and $\beta_{2\text{-em}}$, (ii) contacted the 5′ end of subsite R in the right half of snapshot $\beta_{1\text{-em}}$, and (iii) was at similar distances from O4 of T_{18} and T_{19} (Figure 3), which shows that this $N^{\delta2}$ was situated between the 5′ ends of subsites L and R of 5H-LR in the right halves of snapshots $\alpha_{1\text{-em}}$ and $\alpha_{2\text{-em}}$. These results suggest that $N^{\delta2}$ of Asn235 may be mobile along 5H-LR.

Additionally, we examined the following distances between backbone atoms of Asn235 of BR_B and T_{19} of the C/EBP site in $\alpha_{1\text{-em}}$ versus $\alpha_{2\text{-em}}$: (i) distances from N of the Asn235 backbone to C of the T_{19} backbone were 11.00 Å versus 11.46 Å, and (ii) distances from the same N atom to P of the T_{19} backbone were 12.07 Å versus 12.84 Å. We also found that distances between the centers of mass of Asn235 and T_{19} were 9.52 Å in $\alpha_{1\text{-em}}$ versus 9.94 Å in $\alpha_{2\text{-em}}$. Distance variations indicate displacements of the backbone and centers of mass of Asn235 against T_{19}. These results suggest that Asn235 may be mobile along 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 $\alpha_{1\text{-em}}$ and $\alpha_{2\text{-em}}$ (GCN4 bZIP complex with C/EBP) and between snapshots $\beta_{1\text{-em}}$ and $\beta_{2\text{-em}}$ (GCN4 bZIP complex with C/EBP-1): the left halves of $\alpha_{1\text{-em}}$ and $\alpha_{2\text{-em}}$ differ by four H-bonds and their right halves by eight H-bonds (Table 2); the left halves of $\beta_{1\text{-em}}$ and $\beta_{2\text{-em}}$ differ by 12

Table 2. Different H-Bonding in $\alpha_{1\text{-em}}$ versus $\alpha_{2\text{-em}}$[a]

	basic region[b]		DNA[c]		H-bonding	
					$\alpha_{1\text{-em}}$	$\alpha_{2\text{-em}}$
BR_A	Arg245	$N^{\eta1}$	A3	OP1	no	yes
			A3	O5′	no	yes
	Arg234	$N^{\eta1}$	A2	OP2	no	yes
		$N^{\eta2}$	G1	OP1	no	yes
BR_B	Arg245	$N^{\eta2}$	T18	OP2	no	yes
			T18	O5′	no	yes
	Arg243	$N^{\eta2}$	G8	N7	no	yes
	Arg241	$N^{\eta2}$	T18	OP1	yes	no
	Arg240	$N^{\eta2}$	C7	OP1	no	yes
	Arg234	$N^{\eta1}$	C16	OP1	no	yes
			C16	OP2	no	yes
	Arg232	$N^{\eta1}$	C9	OP1	no	yes

[a]Atom locants follow IUPAC nomenclature (see section S1 of the Supporting Information for transcription). [b]Residues are numbered according to GCN4 (Figure 1). BR_A and BR_B are left and right basic regions in each complex, respectively. [c]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.[4−6] Moreover, we compared direct H-bonds at the GCN4−DNA interface in crystal structures 1YSA and 1DGC[4,5] versus those in their energy-minimized complexes $1YSA_{em}$ and $1DGC_{em}$, 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 $\alpha_{1\text{-em}}$ versus $\alpha_{2\text{-em}}$, and in $\beta_{1\text{-em}}$ versus $\beta_{2\text{-em}}$ (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.[25] Riggs et al. in 1970 reported that Lac repressor accomplishes target-site search ∼2 orders of magnitude faster than random diffusion.[26] In 1981, Berg et al. developed mathematical descriptions of four mechanisms for rapid protein translocation along DNA: sliding, hopping, jumping, and intersegment transfer.[10,11,27−29] 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[30] (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;[31] e.g., sliding of protein Ku was shown by EMSA.[32] The hopping mechanism has also been explored using diverse proteins and captured by various techniques; e.g., hopping of the HoxD9 homeodomain was shown by nuclear magnetic resonance (NMR).[33]

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 HoxD9 homeodomain indicated that DNA-binding domains employ similar structures to both search for and bind to target sites.[33] 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.[7] 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.[35] 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.[34] 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.[36,37] Similarly, by extending the DNA duplex length from 36 to 50000 bp, Khoury et al. observed 15-fold affinity increases for Lac repressor.[38] 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[8] (Table 1); similar results were obtained when we analyzed above the K_d 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.[39] 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.[27] In 5H-LR, the subsite positions differ by only 1 bp. This suggests that the basic region is likely to slide between subsites at 5H-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 5H-LR half-site. We investigated the interface between the basic region and 5H-LR by analyzing snapshots of the interface generated by AMBER simulation. We analyzed K_d values of wt bZIP complexes with target sites that contain 5H-LR versus either subsite; we analyzed K_d differences to investigate how the L and R subsites contribute to affinity between the basic region and 5H-LR. The in vitro and in silico results offer the following insights into how the basic region interacts with 5H-LR.

Our results suggest that the basic region does not recognize 5H-LR solely as a single target site but that it can recognize subsites individually as distinct targets at 5H-LR. The basic region may dissociate from one subsite and reassociate with the other at 5H-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 5H-LR.

Our results also show that when one basic region translocates along 5H-LR, the partner basic region can engage in various DNA-binding activities: (i) at C/EBP, the partner basic region also interacts with 5H-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 5H-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 5H-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,[40−43] 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 5H-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.[9] 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 basic region to translocate between subsites. Therefore, the GCN4 bZIP dimer may use this highly dynamic model to tolerate differences in half-site spacing: e.g., in C/EBP, the two L subsites abut each other, a L subsite and a R subsite from the other copy of 5H-LR overlap by 1 bp, and the two R subsites overlap by 2 bp.

In addition, the EPR results are consistent with solution NMR studies performed on GCN4 bZIP, which demonstrate that the basic region is highly dynamic.[44,45] 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.[4−6] 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.[45] 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 flickering protein secondary structures and high protein mobility upon DNA binding. The authors suggested that fuzziness may be common for proteins binding to specific DNA sites.[46] Similarly, Struhl and co-workers noted that an α-helical bZIP motif displays more structural flexibility than compact globular motifs do, and this allows a bZIP protein to bind different target sites, as described in their work with C/EBP and GCN4 bZIP derivatives.[47] The authors also observed that bZIP proteins contain highly conserved residues responsible for formation of sequence-specific bZIP−DNA complexes, yet bZIP proteins vary widely in their DNA sequence specificities. Johnson examined DNA recognition by GCN4-C/EBP basic region hybrids and also found conformational adaptability in DNA recognition by bZIP proteins, because several hybrids were shown to bind to both the AP-1 and C/EBP sites.[48] Thus, the bZIP is a flexible and versatile motif for the recognition of specific DNA sequences.

Our research shows a case of the basic region interacting with a half-site whose length is increased from 4 to 5 bp. This differs from the GCN4 basic region recognizing a 4 bp cognate half-site as a single target. The benefit of the basic region interacting with a longer half-site is the increase in sequence selectivity during its interaction with DNA. Remarkably, although the basic region interacts with a longer half-site, this does not lead to a longer full-site. CRE and cognate AP-1 are 8 and 7 bp in length, respectively, and the cognate half-sites are 4 bp in length. In contrast, the 5H-LR half-site is 5 bp in length. The basic region interacts with 5H-LR in 7 bp C/EBP-1 and 8 bp AC, C/EBP, and XRE1. The consistent length of the full-sites indicates that although the basic regions may interact with 5H-LR in a highly dynamic manner, the bZIP dimer localizes at the full-site. Therefore, if bZIP transcription factors use this highly dynamic DNA-binding model to interact with their gene regulatory sequence in vivo, bZIP transcription factors still can remain spatially functional at the DNA promoter.

Transcription factors in vivo use their DNA-binding domains to search for and localize to their cognate gene regulatory sequences to govern gene expression. During search, interactions between DNA-binding domains and genomic DNA are dominated by nonspecific interactions, including those between the negatively charged DNA phosphodiester backbone and positively charged protein side chains. As transcription factors reach their gene regulatory sequences, the protein−DNA interactions switch to sequence-specific interactions, which involve base-specific contacts, including H-bonds and van der Waals forces. As discussed above, DNA-binding proteins can use a similar structure, and therefore the same protein surface and DNA-binding residues, to search for and bind to target sites.[7,33] Therefore, the switch between target-site search and localization must involve transformation of the protein−DNA interface: breaking Coulombic, non-specific interactions and establishing base-specific H-bonds for sequence-specific interactions. During this process, sequence-selective (subspecific) interactions must take place; therefore, examination of such interactions can improve our understanding of how transcription factors execute the transition between the search and localization tasks in vivo. Here, we investigated such interactions between the GCN4 basic region and noncognate target sites.

DNA-binding proteins have been shown to use sliding along flanking DNA segments as a mechanism to find their target sites, as discussed above (for a review, see ref 34); these proteins have also been found to dissociate from their target

sites by first sliding onto flanking DNA segments.[35] 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 that 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 subsites, 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.

Notes

The authors declare no competing financial interests.

ABBREVIATIONS

5H-LR, 5 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_A, basic regions in the left half of the bZIP−DNA complex; BR_B, 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-Oriented 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; IHF, 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-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; $\Delta\theta_{app}$, net bound DNA fraction.

REFERENCES

(1) Marcovitz, A., and Levy, Y. (2011) Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc. Natl. Acad. Sci. U.S.A. 108*, 17957−17962.

(2) Hurst, H. C. (1994) Transcription factors 1: bZIP proteins—introduction. *Protein Profile 1*, 123−168.

(3) Agre, P., Johnson, P. F., and McKnight, S. L. (1989) Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. *Science 246*, 922−926.

(4) 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.

(5) 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.

(6) 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.

(7) 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.

(8) Chan, I.-S., Shahravan, S. H., Fedorova, A. V., and Shin, J. A. (2008) The bZIP targets overlapping DNA subsites within a half-site resulting in increased binding affinities. *Biochemistry 47*, 9646−9652.

(9) 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.

(10) von Hippel, P. H., and Berg, O. G. (1989) Facilitated targets location in biological systems. *J. Biol. Chem.* 264, 675−678.

(11) Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20, 6929−6948.

(12) Laurence, T. A., Kwon, Y., Johnson, A., Hollars, C. W., O'Donnell, M., Camarero, J. A., and Barsky, D. (2008) Motion of a DNA sliding clamp observed by single molecule fluorescence spectroscopy. *J. Biol. Chem.* 283, 22895−22906.

(13) Sanchez, H., Suzuki, Y., Yokokawa, M., Takeyasu, K., and Wyman, C. (2011) Protein-DNA interactions in high speed AFM: Single molecule diffusion analysis of human RAD54. *Integr. Biol.* 3, 1127−1134.

(14) 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.

(15) Bird, G. H., Lajmi, A. R., and Shin, J. A. (2002) Sequence-specific recognition of DNA by hydrophobic, alanine-rich mutants of the basic region/leucine zipper motif investigated by fluorescence anisotropy. *Biopolymers* 65, 10−20.

(16) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712−725.

(17) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926−935.

(18) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126−2132.

(19) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics 14*, 33−38.

(20) Lajmi, A. R., Wallace, T. R., and Shin, J. A. (2000) Short, hydrophobic, alanine-based proteins based on the basic region/leucine zipper protein motif: Overcoming inclusion body formation and protein aggregation during overexpression, purification, and renaturation. *Protein Expression Purif.* 18, 394−403.

(21) 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.

(22) Hill, D. E., Hope, I. A., Macke, J. P., and Struhl, K. (1986) Saturation mutagenesis of the yeast *his*3 regulatory site: Requirements for transcriptional induction and for binding by GCN4 activator protein. *Science 234*, 451−457.

(23) Oliphant, A. R., Brandl, C. J., and Struhl, K. (1989) Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: Analysis of yeast GCN4 protein. *Mol. Cell. Biol. 9*, 2944−2949.

(24) 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.

(25) Adam, G., and Delbruck, M. (1968) Reduction of Dimensionality in Biological Diffusion Processes. In *Structural Chemistry and Molecular Biology* (Rich, A., and Davidson, N., Eds.) pp 198−215, W. H. Freeman and Co., San Francisco.

(26) Riggs, A. D., Bourgeois, S., and Cohn, M. (1970) Lac Repressor-operator interaction. 3. Kinetic studies. *J. Mol. Biol. 53*, 401−417.

(27) Winter, R. B., Berg, O. G., and von Hippel, P. H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: Kinetic measurements and conclusions. *Biochemistry 20*, 6961−6977.

(28) Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res. 32*, 3040−3052.

(29) 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.

(30) Zakrzewska, K., and Lavery, R. (2012) Towards a molecular view of transcriptional control. *Curr. Opin. Struct. Biol. 22*, 160−167.

(31) Barsky, D., Laurence, T. A., and Venclovas, C. (2011) How Proteins Slide on DNA. In *Biophysics of DNA-Protein Interactions: From Single Molecules to Biological Systems* (Williams, M. C., and Maher, L. J., Eds.) pp 39−68, Springer, New York.

(32) Paillard, S., and Strauss, F. (1991) Analysis of the mechanism of interaction of simian Ku protein with DNA. *Nucleic Acids Res. 19*, 5619−5624.

(33) Iwahara, J., Zweckstetter, M., and Clore, G. M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A. 103*, 15062−15067.

(34) Shimamoto, N. (1999) One-dimensional diffusion of proteins along DNA: Its biological and chemical significance revealed by single-molecule measurements. *J. Biol. Chem. 274*, 15293−15296.

(35) Kabata, H., Kurosawa, O., Arai, I., Washizu, M., Margarson, S. A., Glass, R. E., and Shimamoto, N. (1993) Visualization of single molecules of RNA polymerase sliding along DNA. *Science 262*, 1561−1563.

(36) Surby, M. A., and Reich, N. O. (1996) Facilitated diffusion of the EcoRI DNA methyltransferase is described by a novel mechanism. *Biochemistry 35*, 2209−2217.

(37) Surby, M. A., and Reich, N. O. (1996) Contribution of facilitated diffusion and processive catalysis to enzyme efficiency: Implications for the EcoRI restriction-modification system. *Biochemistry 35*, 2201−2208.

(38) Khoury, A. M., Lee, H. J., Lillis, M., and Lu, P. (1990) Lac repressor-operator interaction: DNA length dependence. *Biochim. Biophys. Acta 1087*, 55−60.

(39) Wunderlich, Z., and Mirny, L. A. (2008) Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Res. 36*, 3570−3578.

(40) Cranz, S., Berger, C., Baici, A., Jelesarov, I., and Bosshard, H. R. (2004) Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site. *Biochemistry 43*, 718−727.

(41) Berger, C., Piubelli, L., Haditsch, U., and Bosshard, H. R. (1998) Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor. *FEBS Lett. 425*, 14−18.

(42) Kohler, J. J., Metallo, S. J., Schneider, T. L., and Schepartz, A. (1999) DNA specificity enhanced by sequential binding of protein monomers. *Proc. Natl. Acad. Sci. U.S.A. 96*, 11735−11739.

(43) Kohler, J. J., and Schepartz, A. (2001) Kinetic studies of Fos-Jun-DNA complex formation: DNA binding prior to dimerization. *Biochemistry 40*, 130−142.

(44) Saudek, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R., and Pastore, A. (1991) Solution structure of the DNA-binding domain of the yeast transcriptional activator protein GCN4. *Biochemistry 30*, 1310−1317.

(45) 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.

(46) Aguado-Llera, D., Goormaghtigh, E., de Geest, N., Quan, X.-J., Prieto, A., Hassan, B. A., Gómez, J., and Neira, J. L. (2010) The basic helix-loop-helix region of human neurogenin 1 is a monomeric natively unfolded protein which forms a "fuzzy" complex upon DNA binding. *Biochemistry 49*, 1577−1589.

(47) Kim, J., Tzamarias, D., Ellenberger, T., Harrison, S. C., and Struhl, K. (1993) Adaptability at the protein-DNA interface is an important aspect of sequence recognition by bZIP proteins. *Proc. Natl. Acad. Sci. U.S.A. 90*, 4513−4517.

(48) Johnson, P. F. (1993) Identification of C/EBP basic region residues involved in DNA sequence recognition and half-site spacing preference. *Mol. Cell. Biol. 13*, 6919−6930.

(49) Hope, I. A., and Struhl, K. (1985) GCN4 protein, synthesized in vitro, binds HIS3 regulatory sequences: Implications for general control of amino acid biosynthetic genes in yeast. *Cell 43*, 177−188.

(50) Bird, G. H., and Shin, J. A. (2002) MALDI-TOF mass spectrometry characterization of recombinant hydrophobic mutants

containing the GCN4 basic region/leucine zipper motif. *Biochim. Biophys. Acta 1597*, 252−259.