*Plant Molecular Biology* **41:** 1–13, 1999.



# Bipartite determinants of DNA-binding specificity of plant basic leucine zipper proteins

Xiping Niu[1], Lisa Renshaw-Gegg[2], Linda Miller[3] and Mark J. Guiltinan[3,*]
[1]*Pioneer Hi-Bred International, Inc., 7300 NW 62nd Ave., Johnston, IA 50131;* [2]*Amgen Inc, 14-2-C Amgen center, Thousand Oaks, CA 91230;* [3]*Department of Horticulture, 306 Wartik Laboratory, Biotechnology Institute, Pennsylvania State University, University Park, PA 16802, USA (*author for correspondence)*

Received 23 April 1998; accepted in revised form 8 April 1999

*Key words:* bZIP, EmBP-1, TGA1a, mutagenesis, DNA binding specificity

**Abstract**

The basic leucine zipper (bZIP) proteins are one of the largest and most conserved groups of eukaryotic transcription factors/repressors. Two major subgroups among the plant bZIP proteins have been identified as G-box (CCACGTGG) or C-box (TGACGTCA) binding proteins based on their DNA binding specificity and the amino acid sequences of their basic regions. We have investigated how plant bZIP proteins determine their DNA binding specificity by mutation of the basic domain of the G-box-binding protein EmBP-1. Four subregions of the EmBP-1 basic domain that differ from the C-box-binding protein TGA1a were substituted singly or in combination with the corresponding regions of TGA1a. DNA binding experiments with the mutant proteins demonstrated that binding specificity of plant bZIP proteins is determined independently by two regions, the core basic region and the hinge region. These two regions have an additive effect on DNA binding specificity. PCR-assisted binding-site selections using key mutants demonstrated that only G-box and C-box binding specificity can be generated by combinations of amino acids in the basic domains of EmBP-1 and TGA1a. These results suggest that factorial contributions of the amino acid residues in the basic domain combine to determine DNA-binding specificity of bZIP proteins.

## Introduction

Many studies have revealed vital roles of basic leucine zipper (bZIP) proteins in biological processes in eukaryotic cells. In plants, the bZIP proteins are involved in control of seed storage gene expression, photomorphogenesis, and organ establishment (Schmidt *et al.*, 1987, 1990; Oyama *et al.*, 1997; Walsh *et al.*, 1998). Furthermore, the plant bZIP proteins are also implicated in regulating gene expression in response to abscisic acid (ABA), light, anaerobiosis, and developmental signals (Menkens *et al.*, 1995). Individual plants can express a large number of bZIP proteins; for example, in *Arabidopsis*, 13 bZIP proteins have been identified (Figure 1). It is likely that differential expression of bZIP genes, hetero-dimerization and unique binding specificity of the different proteins result in a factorial mechanism generating functional diversity. Thus, it is essential to the understanding of plant gene regulation to determine the molecular mechanisms governing DNA binding specificity of the plant bZIP proteins.

Amino acid sequence alignment of the basic regions of plant bZIP proteins indicates that they are more related to each other than to those from other eukaryotes, indicating they are an ancient evolutionary subgroup of bZIP proteins (Johnson, 1993). DNA binding studies have shown that the binding specificity of plant bZIP proteins are related by the nearly universal recognition of an ACGT core sequence (Foster *et al.*, 1994). The binding affinity of a particular bZIP protein is largely determined by the three bases flanking the four core nucleotides, although binding specificity is variably flexible.

Among plant bZIP proteins, three major groups can be distinguished based on DNA binding specificity. Some prefer the G-box (−3C −2C −1A −0C G+0 T+1 G+2 G+3), others the C-box (−3T −2G

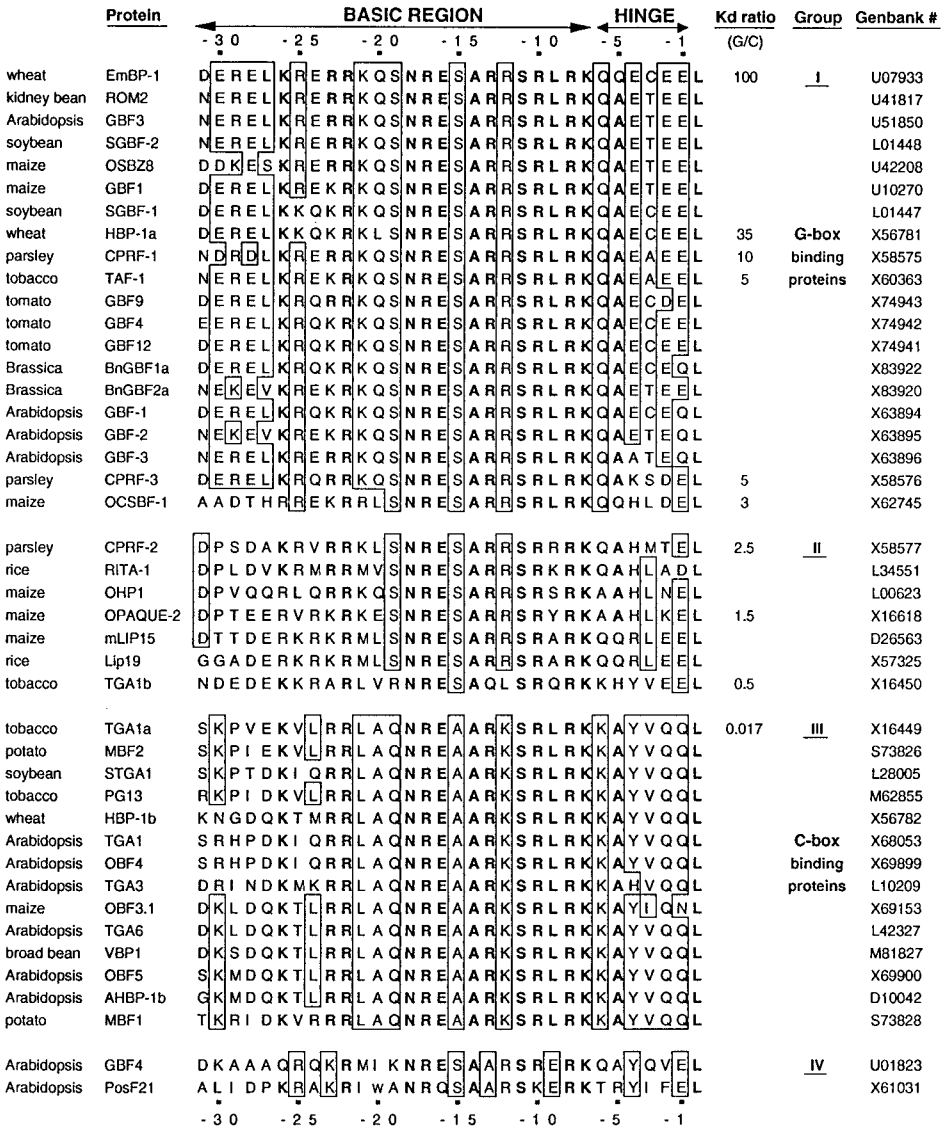


*Figure 1.* Alignment of the basic and hinge regions of plant bZIP proteins. The plant bZIP proteins are classified as four groups according to their DNA binding specificity and amino acid sequences in the basic-hinge regions. Groups I–III are named according to Izawa *et al.* (1993). Group IV proteins do not bind to DNA as homodimers. Conserved amino acids among the plant bZIP proteins are in bold type. The amino acids that are conserved in each group are boxed. Amino acid numbering is according to Suckow *et al.* (1993). The first leucine in leucine heptad repeats is numbered +1, the last amino acid of the basic and hinge regions is −1. The ratio of G-box to C-box dissociation constants ($K_d$) of ten plant bZIP proteins was determined by Izawa *et al.* (1993). GenBank accession numbers for these bZIP proteins are listed. The basic-hinge region of yeast bZIP protein GCN4 is also shown at the bottom for comparison with plant bZIP proteins.

−1A −0C G+0 T+1 C+2 A+3), and the rest bind G-box or C-box with about the same affinity (Izawa *et al.*, 1993). EmBP-1 is the strongest and most specific G-box-binding protein among 10 plant bZIP proteins tested with about 100-fold difference in $K_d$ ratio of G-box vs. C-box-binding affinity, whereas TGA1a is the strongest and most specific C-box-binding protein with a value of 0.017 in $K_d$ ratio of G-box vs. C-box binding affinity (Izawa *et al.*, 1993). EmBP-1 was isolated from wheat by interaction with an ABA response element (Guiltinan *et al.*, 1990). Its DNA binding specificity and DNA binding and dimerization domains have been characterized, indicating that EmBP-1 is a typical bZIP transcription factor which binds to a G-box site (GCCACGTGGC) with highest affinity (Guiltinan and Miller, 1994; Niu and Guilti-

nan, 1994). TGA1a was cloned from tobacco and originally shown to bind the sequence TGACG in the as-1 element of the 35S promoter of the cauliflower mosaic virus (Katagiri *et al.*, 1989). To investigate the determinants of DNA binding specificity, we measured G-box and C-box binding by EmBP-1 mutants with TGA1a amino acid substitutions. Our results indicate DNA-binding specificity of plant bZIP proteins is determined not only by the basic region but also by the hinge region. The results also illustrate that combinatorial contribution to DNA binding specificity is important in determining high DNA-binding specificity of the bZIP proteins.

## Materials and methods

### *PCR-assisted mutagenesis*

Mutations were introduced into EmBP-1 coding sequence for the basic-hinge region by a PCR-assisted mutagenesis method (Ho *et al.*, 1989), using the plasmid pXN11 (Niu and Guiltinan, 1994) as template. The PCR products were cut with *Bam*HI and *Hin*dIII and cloned into similarly digested pXN11.

### *PCR-based transcription/translation*

EmBP-1 and its mutant proteins were synthesized *in vitro* in transcription/translation reactions as described (Guiltinan and Miller, 1994), using PCR products amplified from the above plasmids containing the sequences of EmBP-1 and its mutants. In PCR reactions, the N-terminal primer with a phage T7 promoter sequence starts from amino acid 245 of EmBP-1a (5′-GGAATTCTAATACGACTCACTATAGGGAGCCATGTCGCTGTCTCAG-3′), the C-terminal primer ends with amino acid 317 of EmBP-1a as described (Guiltinan and Miller, 1994). EmBP-1 and its mutant proteins synthesized *in vitro* were used in electrophoretic mobility shift assays shown in Figure 2.

### *Electrophoretic mobility shift assay (EMSA)*

EMSA was as previously described (Guiltinan and Miller, 1994). For G-box and C-box probes, two oligonucleotides (G-box: 5′-TGGATGGCCACGTGGCCATCCA-3′; C-box: 5′-AGCTTAGTGACGTCACTAAGCT-3′) were cloned into the *Sal*I-cut and Klenow-filled pUC19 vector, generating plasmids pMG92.23 and pLWRGc-14. pMG92.23 and pLWRGc-14 were digested with *Eco*RI and *Pst*I. The fragments containing G-box and C-box were gel-purified and labeled by Klenow filling to high specific activity (1–2 $\times$ $10^8$ cpm/$\mu$g DNA). Minor differences in labeling were equalized by dilution with 1$\times$ binding buffer. Quantification of mobility shift gels was performed with a PhosphorImager (Molecular Dynamics).

### *Overexpression and purification*

For large-scale purification of recombinant proteins, XL1-blue cells containing the plasmids were grown to $OD_{600}$ 0.4 and induced with isopropyl-1-thio-$\beta$-D-galactopyranoside for 3 h. Proteins were purified as described (Niu and Guiltinan, 1994). The buffer containing the eluted proteins was changed three times with storage buffer (50 mM Tris pH 7.6, 100 mM KCl) by Centricon-30 filter ultrafiltration (Amicon). Proteins were diluted with an equal volume of glycerol and stored at $-20$ °C. Purified recombinant proteins were used in PCR-assisted binding-site selection and DNase I footprinting.

### *PCR-assisted binding-site selection*

PCR-assisted binding-site selection was performed as previously described (Niu and Guiltinan, 1994) with the following modifications: the oligonucleotide with a 14 nucleotide random sequence was made double-stranded DNA and $^{32}$P-labeled in a buffer consisting of 50 mM Tris-HCl pH 7.2, 100 mM NaCl, 10 mM $MgCl_2$, and 0.1 mM DTT. In binding reactions, the concentrations of poly(dI-dC) varied from 50 to 5 ng/$\mu$l depending on individual mutant proteins. The bound fractions were separated from unbound fractions on 4% polyacylamide (29:1 ratio of acrylamide to bisacrylamide)-0.5$\times$ Tris/borate/EDTA gels at room temperature.

### *DNase I footprinting*

DNA probe preparation, footprinting reactions and denaturing gels were as described (Niu *et al.*, 1996). Poly(dI-dC) concentrations vary according to individual mutants as indicated in the text. For some mutants, reaction volumes were scaled up proportionally to larger volumes to accommodate more mutant protein in a reaction. For C-box probe, pLWRGc-14 was used as template for PCR reaction to make DNA probes as previously described (Guiltinan and Miller, 1994; Niu and Guiltinan, 1994). For A-box and AP-1 site probes, cloned DNA binding sites 12 and 13 in Figure 3D

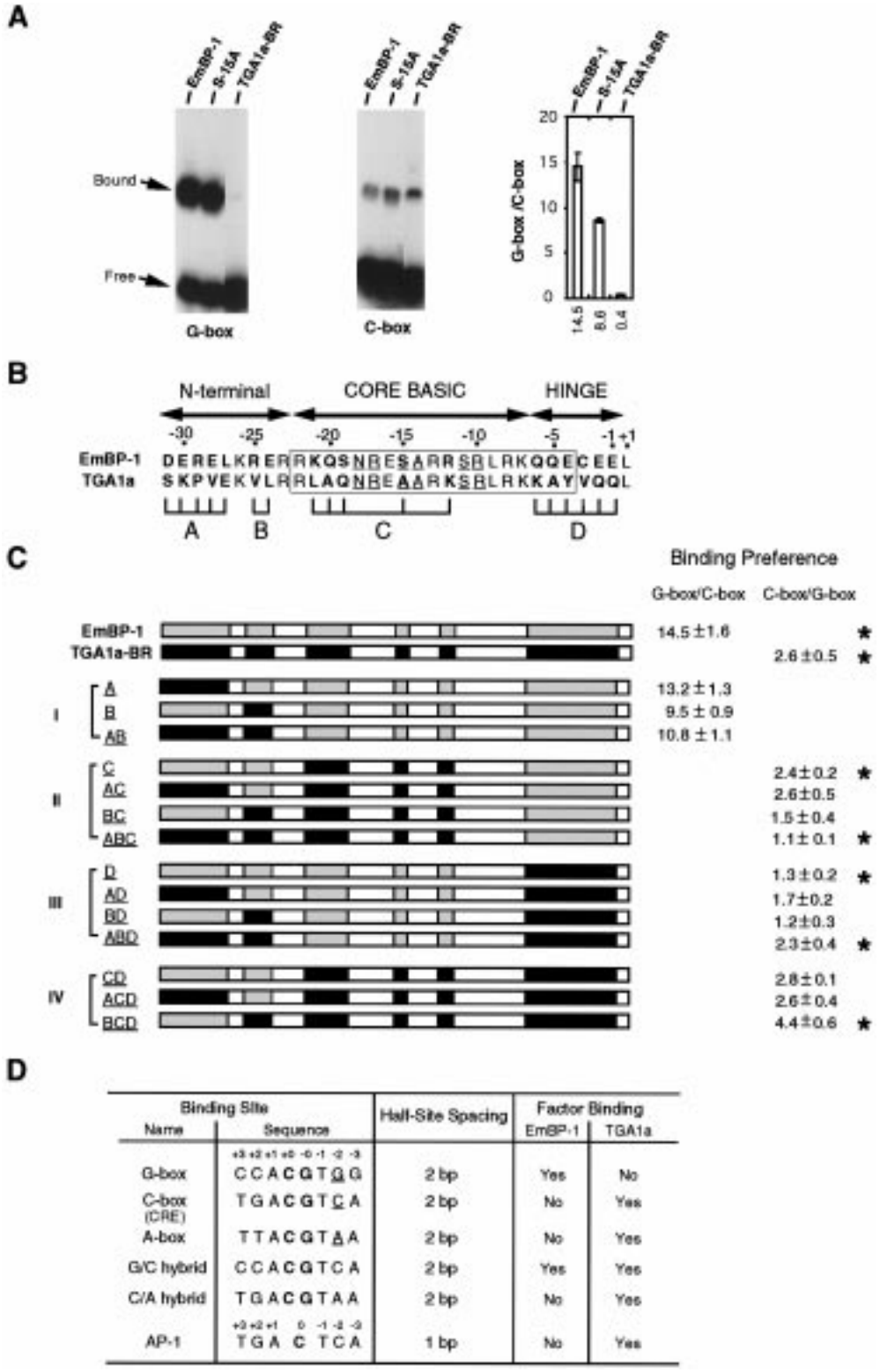
A
EmBP-1
S-15A
TGA1a-BR
Bound
Free
G-box
C-box
G-box /C-box
20
15
10
5
0
14.5
8.6
0.4
B
N-terminal
CORE BASIC
HINGE
-30
-25
-20
-15
-10
-5
-1
+1
EmBP-1 DERELKRERRKQSNBESARRSBLRKQQECEEL
TGA1a SKPVEKVLRRLAQNBEAARKSBLRKKAYVQQL
A
B
C
D
C
Binding Preference
G-box/C-box
C-box/G-box
EmBP-1
TGA1a-BR
14.5 ±1.6
2.6 ±0.5
I
A
B
AB
13.2 ±1.3
9.5 ± 0.9
10.8 ±1.1
II
C
AC
BC
ABC
2.4±0.2
2.6±0.5
1.5 ± 0.4
1.1 ± 0.1
III
D
AD
BD
ABD
1.3 ± 0.2
1.7±0.2
1.2±0.3
2.3±0.4
IV
CD
ACD
BCD
2.8 ± 0.1
2.6 ± 0.4
4.4 ±0.6
D
Binding Site
Name
Sequence
Half-Site Spacing
Factor Binding
EmBP-1
TGA1a
G-box
CCACGTGG
2 bp
Yes
No
C-box (CRE)
TGACGTCA
2 bp
No
Yes
A-box
TTACGTAA
2 bp
No
Yes
G/C hybrid
CCACGTCA
2 bp
Yes
Yes
C/A hybrid
TGACGTAA
2 bp
No
Yes
AP-1
TGA C TCA
1 bp
No
Yes

*Figure 2.* Analysis of EmBP-1 mutants. A. Gel shift assays of S-15A mutant. Left: EmBP-1 and its mutants bind to G-box probe. Middle: EmBP-1 and its mutants bind to C-box probe. Right: quantification of gel shift results. Ratio of bound G-box to C-box probes from three experiments and the standard errors are presented. B. Alignment of the basic-hinge regions of EmBP-1 and TGA1a. Amino acid differences between EmBP-1 and TGA1a are in bold type. The amino acid positions that make base-specific contacts with DNA in GCN4 crystal structures are underlined. The regions that make phosphate backbone contacts with DNA are boxed. Amino acid numbering and region designations are as described in the text and indicated. C. DNA binding of EmBP-1 subregion substitution mutants. Horizontal bars represent the basic-hinge regions of EmBP-1 and its mutants. EmBP-1 specific and TGA1a specific amino acids are represented by gray and black boxes, respectively. Mutant designations and grouping are shown at the left. Binding preferences of mutant proteins are represented at the right by the ratio of bound G-box to C-box. The ratios are the average of three experiments, the standard errors are also shown. The stars indicate that PCR-assisted binding-site selection was performed on these mutants (see Figure 3). D. DNA binding sites bound by EmBP-1 or TGA1a. Except for the AP-1 site, nomenclature of these binding sites follow that of Izawa *et al.* (1993, 1994). The C-box is identical to the CRE site. The sequences of these binding sites are shown with the position −2 underlined to indicate G-box, C-box and A-box signature. Half-site spacing and binding to these sites by EmBP-1 or TGA1a are also indicated.

were used as templates in PCR reactions to make DNA probes, respectively.

## Results

### *Single amino acid substitution at −15 does not alter DNA-binding specificity of EmBP-1*

According to X-ray crystal structures of GCN4 complexed with DNA-binding sites (Ellenberger *et al.*, 1992; König and Richmond, 1993; Keller *et al.*, 1995), base-specific contacts involve amino acids Asn-18, Ala-15, Ala-14, and Arg-10 (nomenclature follows that used by Suckow *et al.* (1993), the first leucine of the heptad repeats is numbered +1, the last amino acid of the basic domain is −1). Ser-11 or Thr-17 is also involved in base-specific contact of GCN4 binding to the AP-1 site (TGACTCA) or CRE site (TGACGTCA). Consistent with the X-ray structures, mutagenesis studies of GCN4 and mammalian bZIP protein C/EBP demonstrated that amino acids from −18 to −11 are major residues determining DNA-binding specificity (Johnson, 1993; Suckow *et al.*, 1993a,b, 1996).

Among these amino acids that make base-specific contacts, a change of Ala-14 to Val-14 brings about the tendency of GCN4 to bind to the C/EBP-binding site (ATTGCGCAAT) (Johnson, 1993), demonstrating that replacement of a single base-contact residue can dramatically alter DNA-binding specificity. Systematic single amino acid substitutions at positions −18, −15, −14, −11, and −10 of GCN4 have also been shown to change GCN4 DNA-binding specificity (Suckow *et al.*, 1993a,b, 1994a,b).

Based on the results with GCN4, we compared residues in plant bZIP proteins predicted to contact DNA. From positions −18 to −11, plant G-box and C-box binding proteins differ in two residues at positions −15 and −12. Ser-15 and Arg-12 are conserved in plant G-box binding proteins, whereas Ala-15 and Lys-12 are conserved in C-box binding proteins (Figure 1). Because Ala-15 makes base-specific contact with DNA in GCN4 crystal structures, it is reasonable to postulate that the amino acid at −15 may be a major determinant for plant G-box and C-box binding specificity (Foster *et al.*, 1994).

To test this hypothesis, Ser-15 in EmBP-1 was replaced by Ala-15 of TGA1a, generating a mutant S-15A. The S-15A mutant was translated *in vitro* and tested by EMSA against G-box and C-box binding sites along with the wild-type EmBP-1 and a TGA1a-like mutant in which the entire basic and hinge regions of EmBP-1 are replaced by those of TGA1a (we designate this mutant as TGA1a-BR). The results in Figure 2A show that S-15A still binds the G-box with high affinity comparable to that of EmBP-1 (left panel in Figure 2A). Additionally, S-15A mutant did not bind to the C-box much more strongly than EmBP-1 (Figure 2A, middle panel). To minimize potential effects of protein concentration and stability on comparison of DNA binding preference, the ratio of bound G-box probe to bound C-box probe for each protein was calculated (Figure 2A, right panel). The ratio for S-15A has a value of 8.6, which is much higher than that of TGA1a-BR (0.4) but lower than that of EmBP-1 (14.5). This result indicates that S-15A has strong preference for G-box, thus this mutation does not change EmBP-1 DNA binding specificity dramatically. This value for EmBP-1 G/C box specificity is about 7 times lower than previously reported (Izawa *et al.*, 1993), the difference most likely being due to different experimental conditions. In the previous report (Izawa *et al.*, 1993) crude protein extracts from *Escherichia coli* were used, whereas in this report, *in vitro* translated proteins were used. How-

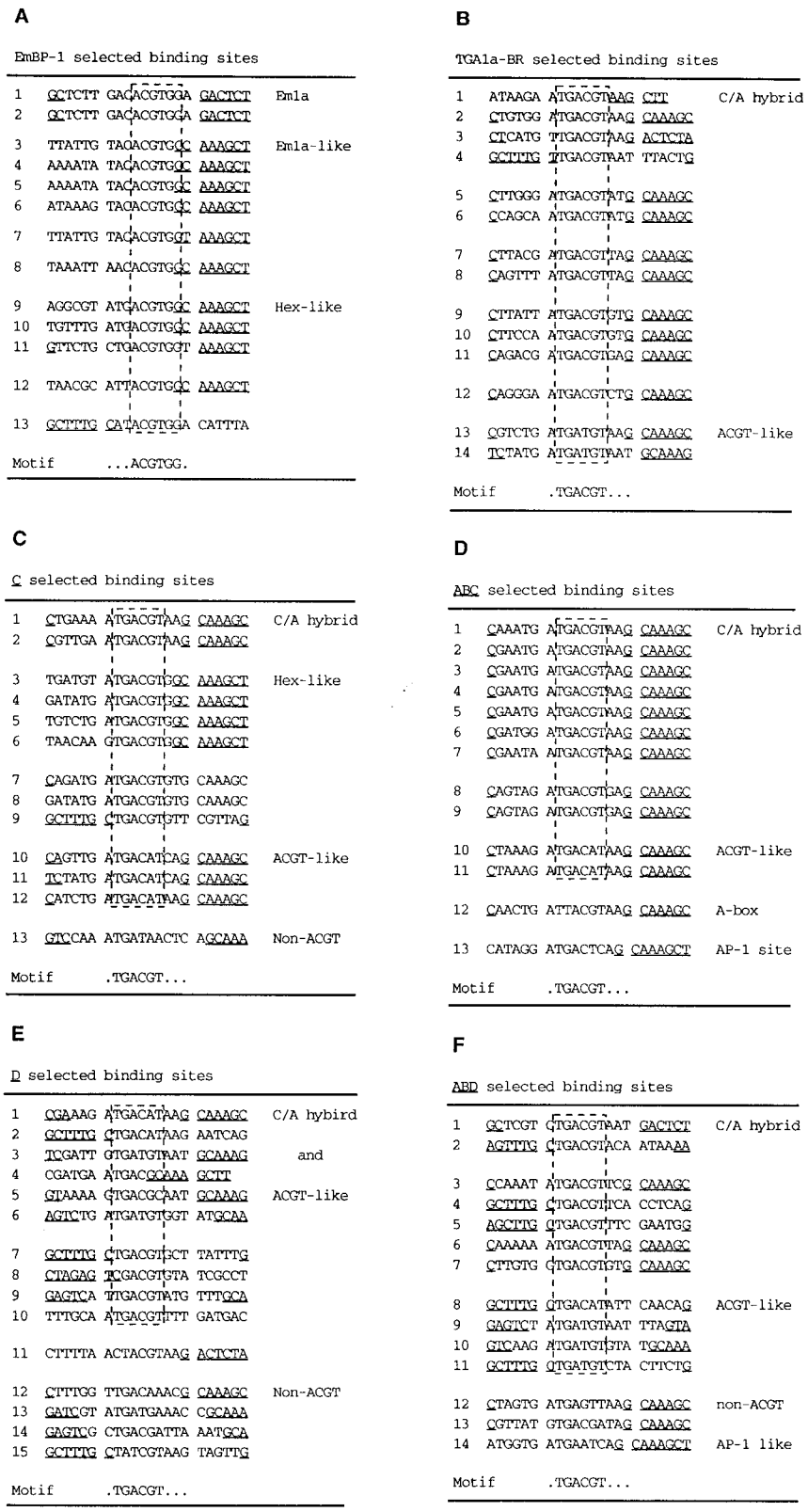


*Figure 3.* Selected binding sites from a random oligonucleotide pool by EmBP-1 and its mutants. A to F are the selected binding sites by EmBP-1 and its mutants (TGA1a-BR, C, ABC, D, and ABD. Selected binding sites (assuming a 10 bp sequence) are aligned centering the ACGT core sequence. Sequence motifs identified in the selected binding site pools are boxed with a dashed line and summarized at the bottom of each pool. Some known DNA binding sites (Em1a, Hex, C/A hybrid, A-box, AP-1 site) contained in the selected binding sites are indicated at the right. The oligonucleotide sequences flanking the 14 bp random sequences are underlined.

ever, qualitatively, these two experimental values are in agreement.

The above results suggest that amino acids besides −15 also contribute to G-box and C-box binding specificity. Because Ser-l5/Ala-15, and Arg-12/Lys-12 belong to two groups with similar biochemical and physical properties (small/neutral and basic, respectively), their relative contributions at each position to G-box and C-box binding specificity may not be large.

*Substitution of amino acids in the basic-hinge regions of EmBP-1 by amino acids in TGA1a*

Because single amino acid substitution does not significantly affect G-box and C-box binding preference of plant bZIP proteins, we investigated what subregions in the basic-hinge regions of plant bZIP proteins may be important for determining G-box and C-box binding specificity. As shown in Figure 2B, the amino acid differences between the basic-hinge regions of EmBP-1 and TGA1a can be divided into four subregions: region A contains five amino acids from −31 to −27, region B contains two amino acids at −25 and −24, region C contains five amino acids at −21, −20, −19, −15, and −12, and region D contains six amino acids from −6 to −1 representing the hinge region. Amino acids in regions A, B, and C are clustered and may act as independent specificity determining groups. Region C consisted of five amino acids that are in a region normally referred to as core basic region. All these regions and permutations in the basic-hinge domain of EmBP-1 were substituted by corresponding amino acids of TGA1a, generating a total of 15 EmBP-1 mutants. For nomenclature, all EmBP-1 mutants will be referred to by the region(s) which have been replaced by the corresponding TGA1a sequence except for TGA1a-BR, in which the entire basic-hinge region of EmBP-1 was replaced by the TGA1a basic-hinge region.

*DNA binding of EmBP-1 mutants*

To assess the DNA binding specificity of all the mutants, mutant proteins as well as EmBP-1 were synthesized by *in vitro* translation using rabbit reticulocyte lysate. Translated products of correct size were verified by SDS protein gels, quantified by Phosphorimager, and adjusted to equal molar concentration. The same amounts of mutant and EmBP-1 proteins were tested against G-box and C-box binding site probes in gel mobility shift assays, which were repeated three times.

Gel mobility retardation experiments indicated that all the mutants were still capable of binding to either G-box or C-box sites in the conditions tested (data not shown). To estimate DNA binding preference of the mutant proteins, ratios of bound G-box probe to bound C-box probe were measured using a Phosphorimager and calculated from three replicate experiments. Figure 2C indicates that the A, B, or AB mutants (Group I mutants) retained G-box binding specificity comparable to EmBP-1. Ratios of bound G-box probe to C-box probe are very high and comparable to that of EmBP-1. This suggests that the A and B regions of plant bZIP proteins exert little effect on DNA-binding specificity. However, this region is an integral part of the basic region and a small deletion completely abolishes DNA binding (Schindler et al., 1992; Guiltinan and Miller, 1994). Figure 2C also indicates that the rest of the EmBP-1 mutants (Groups II, III, IV) change to weak or strong C-box-binding proteins judging by ratios of bound C-box probe to G-box probe. These mutants contain mutations in C or D region or both. Combined mutation of the B, C and D sites gave the highest C-box-binding preference examined.

*Core basic region can determine DNA binding specificity*

X-ray crystal structures of GCN4 indicate that amino acids from −22 to −4 in GCN4 make 5 base-specific and 9 phosphate backbone contacts with its DNA targets (see Figure 2B for these positions in EmBP-1 and TGA1a). Most of these contacts occur within amino acids from −22 to −7 (5 base-specific contacts and 8 DNA backbone contacts). We designate this region as the core basic region. One backbone contact is extended to the region from −6 to −1, adjacent to the leucine zipper region. We designate this region as the hinge region. C and all other Group II mutants have five amino acid substitutions spanning from −21 to −11 (Figure 2B). Because the mutations are flanked by the conserved amino acids of EmBP-1 and TGA1a, the changed region is actually from −23 to −7. Therefore, C mutants change the entire core basic region from EmBP-1 to TGA1a.

Ratios of bound C-box probe vs. G-box probe measured from gel mobility shift assays indicate that the C mutant and its derivatives (Group II) exhihited C-box binding preference (Figure 2C). C and AC have a ratio of bound C-box probe to C-box probe comparable to that of TGA1a-BR mutant. However, the ratio of bound C-box to bound G-box for BC and ABC mu-

tants is low; in particular, ABC has a ratio near 1. Gel shift assays indicate that ABC binding either to G-box or to C-box probes was very weak, making quantification subject to large deviation. One reason for such weak binding to DNA by ABC may be that the protein-DNA complex is not stable. Another possibility is that a new binding specificity was generated which was not represented in the probes used for this experiment.

To further investigate DNA binding specificity of Group II mutants, PCR-assisted binding-site selections were applied to C and ABC mutants. We previously used this method to determine the wild-type binding specificity of EmBP-1 (Niu and Guiltinan, 1994). As controls, EmBP-1 and TGA1a-BR were also subject to binding-site selections under our modified conditions (see Materials and methods). As discussed above, recombinant ABC mutant protein did not bind to DNA strongly, therefore to achieve binding by ABC, we reduced poly(dI-dC) concentrations 5–10-fold in binding reactions. After five rounds of selection, selected DNA sequences were cloned and sequenced as shown in Figure 3A–3D.

The results indicated that EmBP-1 and TGA1a-BR selected binding sites with the G-box and C-box motifs, respectively (Figure 3A and 3B). All of EmBP-1 selected binding sites contain the motif of ACGTGG, which has a high-affinity G-box half site (T+1 G+2 G+3). A representative binding site selected by EmBP-1 is the Em1a site (GACACGTGGA), which was originally used to clone EmBP-1 (Guiltinan *et al.*, 1990). All of TGA1a-BR selected binding sites have a motif of TGACGT (reverse complement sequence is ACGTCA), which contains a typical strong C-box half site ($-3$ T $-2$ G $-1$ A). In conjunction with this strong C-box half site, the second half site can be variable but they are not random sequences, as has been shown for Arabidopsis TGA1 (Schindler *et al.*, 1992), a homologue of tobacco TGA1a. One of commonly found secondary half sites is T+1 A+3 A+4. A combination of these two half sites, $-1$T $-2$G $-3$A and T+1 A+3 A+4, is the C/A hybrid site, following previous nomenclature (Izawa *et al.*, 1993, 1994), which is also selected by TGA1a-BR.

Binding sites selected by C and ABC are shown in Figure 3C and 3D. Both C and ABC selected binding sites are C-box-like with a clear sequence motif of $-3$T $-2$G $-1$A $-0$C G+0 T+1. The C/A hybrid site is also in both pools of selected binding sites. In the former sequence pool, C selected the Hex sequence (G/C hybrid site) (Tabata *et al.*, 1989), which is bound by both plant G-box and C-box proteins (Tabata *et al.*, 1991; Schindler *et al.*, 1992a, b). The Hex sequence was also selected by EmBP-1 (Figure 3A). Furthermore, all of the mutants we generated bound to the Hex sequence (data not shown). In the latter sequence pool, ABC also selected an A-box (TTACGTAA) sequence and the AP-1 (TGACTCA) site (Figure 3D). Both of these two sequences have been shown to be bound by TGA1a (Izawa *et al.*, 1993; de Pater *et al.*, 1994).

As shown in our previous study of wild-type EmBP-1 (Niu and Guiltinan, 1994), selected binding sites are a pool of DNA sequences of varying affinities. In this study, gel shift assays and PCR binding-site selections demonstrated that C and ABC mutants are C-box-binding proteins. A-box and AP-1 sites have been reported to be bound by C-box-binding proteins and are in the pool of ABC selected DNA sites. To verify DNA binding specificity and to further examine relative affinity of mutant proteins for these binding sites, we tested C and ABC binding to these three binding sites by footprint analysis.

As shown in Figure 4, TGA1a-BR binds to C-box and A-box strongly with slightly higher affinity for C-box (Figure 4, lanes 1–4). C mutant also binds to C-box and A-box strongly with slightly higher affinity for A-box (Figure 4, lanes 5–8). Binding of ABC mutant to these two sites is much weaker but protection of the binding sites can still be seen (lanes 9–12). Under the same protein concentrations used for C-box and A-box binding, footprinting titrations indicate all these mutants can bind to the AP-1 site with sharply reduced affinity. None of these mutants showed specific binding to the G-box as indicated by DNase I footprinting and gel mobility shift assays (data not shown).

*The hinge region influences DNA-binding specificity*

The hinge region defines the junction between the dimerization interface and the DNA-binding domain. In D mutants, the hinge region (amino acids $-6$ to $-1$) of EmBP-1 is replaced by that of TGA1a which differs at all 6 positions. As shown in Figure 2C, the D mutant and its derivatives AD, BD, and ABD (Group III) also prefer C-box to G-box as indicated by ratios of bound C-box to bound G-box probes. Except the ABD mutant, however, preference for C-box by Group III mutants is not strong because these mutants have a ratio of slightly higher than 1. The possible reasons for this are that D mutants may not retain stringent binding specificity or that a new binding specificity is generated. Like the ABC mutant, Group III mutants also

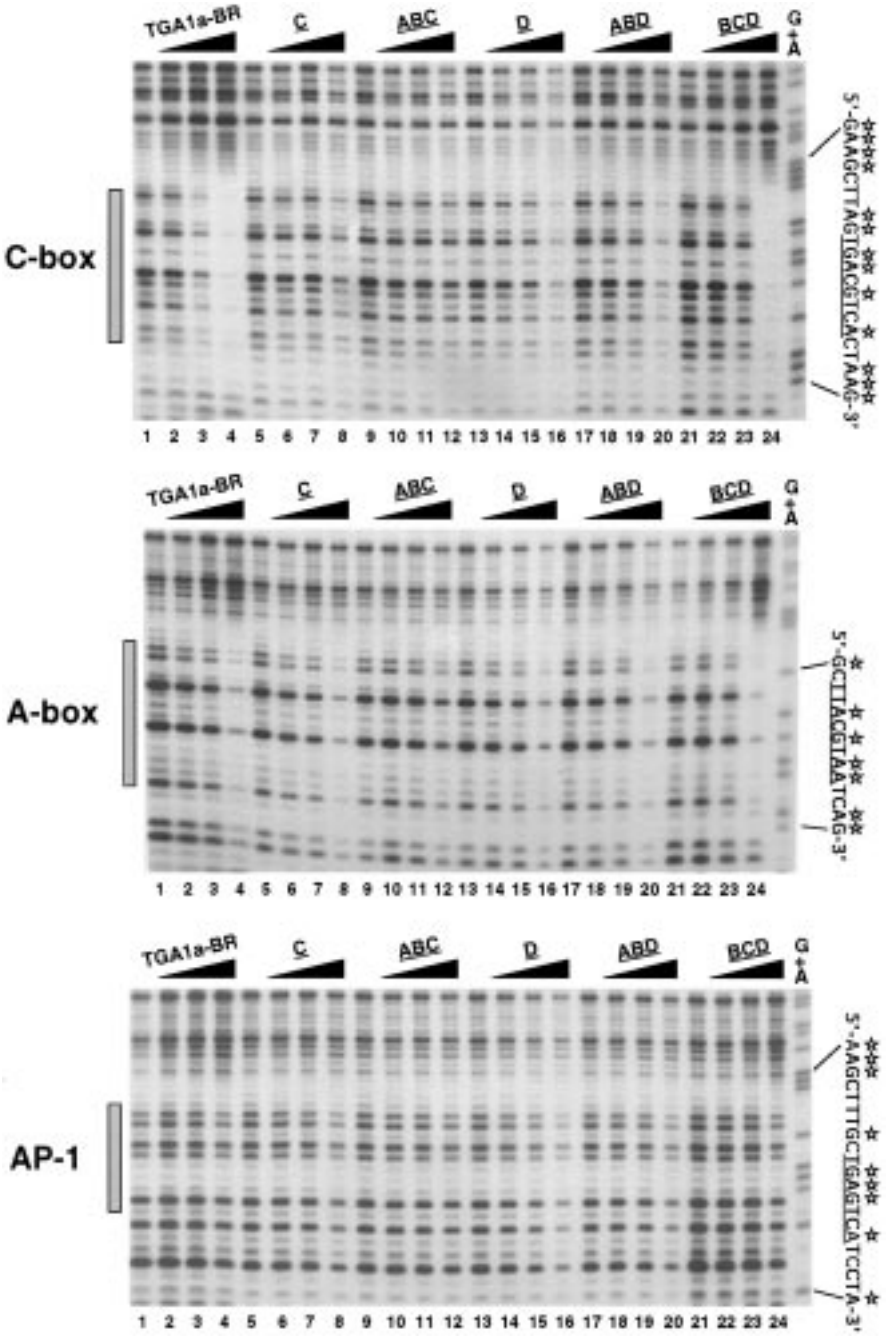


*Figure 4.* DNase I footprinting of key EmBP-1 mutant proteins. From top to bottom are EmBP-1 mutant proteins binding to C-box, A-box, and AP-1 sites. Each mutant protein was tested with the protein concentration in 3-fold increments (1×, 3×, and 9×). For each mutant, the same amount of protein was tested against three different binding sites. Vertical bars at the left indicate bases protected by mutant proteins. G+A ladders are shown at the right. Bottom strand sequences of the DNA probes are also shown with G and A residues indicated by stars. The underlined DNA sequences indicate C-box, A-box and AP-1 site, respectively.

seem not to form very stable complexes with DNA as indicated by gel shift assays. These results suggest that swapping of the hinge regions between G-box- and C-box-binding proteins resulted in reduced DNA-binding affinity of mutant proteins. Therefore, some D mutants as well as C mutants do not bind to DNA as avidly as EmBP-1 or the CD mutants.

To further examine the binding specificity of D mutants, PCR-assisted binding-site selections were applied to D and ABD mutants. In the binding reactions for site selection by recombinant D and ABD proteins, less poly(dI-dC) was necessary to achieve binding occupancy similar to EmBP-1 and TGA1a-BR (data not shown). This result is consistent with the use of *in vitro* translated proteins for our initial gel shift assays. After five rounds of selection, over 14 individual DNA binding sites for each mutant were cloned and sequenced. As shown in Figure 3E and 3F, most of the D and ABD selected binding sites have ACGT or ACGT-like core sequences. TGACGT is still the major motif among the selected binding sites for both mutants (boxed with dashed line in Figure 3E and 3F). T+1 A+2 A+3 half site has also been selected in combination with the C-box half site −3T −2G −1A (C/A hybrid in Figure 3E and 3F). These results demonstrated that D and ABD mutants are C-box-binding proteins. However, comparison of selected binding sites by D and ABD with those by C, ABC, and TGA1a-BR indicate that the D mutant consensus sequences are not as uniform as the C, ABC and TGA1a-BR selected sites. First, there are many more ACGT-like sequences (−0C is substituted by T, G+0 by A, or T+1 by C) in the pools of D and ABD mutant selected binding sites. Second, half sites that associated with the −3T −2G −1A half site are more random. Third, there are more sequences without an ACGT core (non-ACGT sequences), especially in D selected binding sites.

To verify binding specificity of D mutants, DNase I footprinting was performed on D and ABD mutants against C-box, A-box, and AP-1 sites. As shown in Figure 4 (lanes 13–16 and lanes 17–20), these mutants do bind to these binding sites specifically. However, the D mutant has a relatively relaxed binding specificity as indicated by protection of entire DNA probe sequences on the footprinting gels (Figure 4, lanes 13–16). A more stringent C-box-binding specificity was observed in the ABD mutant, in which mutations in regions A and B of the basic region were added to the D mutation (Figure 4, lanes 17–20). Binding affinities by D mutants for C-box and A-box are approximately the same, but the affinity for the AP-1 site is reduced. Because TGA1a protein was shown to bind the AP-1 site, D mutants are good candidates to investigate if this is due to the hinge region. However, we have not detected a stronger preference of these mutants for the AP-1 site compared to C mutants. These results were also verified by gel shift assays (data not shown).

*Additive effect of core basic and hinge regions*

The results above indicate that the regions most important in determining DNA-binding specificity are the C and D regions. The effects of C and D mutations appear to be additive because the CD mutants (Group IV and TGA1a-BR in Figure 2C) showed stronger C-box-binding activity than either C mutants (Group II) or D mutants (Group III) alone. Statistical analysis of these data using Student's $t$-test showed that these values are significantly different at at least the 95% confidence level. Figure 2C indicates that the ratio of bound C-box to bound G-box probes for CD mutants is overall higher than that of either C mutants or D mutants. PCR-assisted binding-site selection by the BCD mutant showed that BCD selected binding sites are similar to those selected by the strong C-box-binding proteins TGA1a-BR (data not shown). DNase I footprinting in Figure 4 (lanes 21–24) demonstrates that BCD binds to the C-box strongly but to the A-box with slightly lower affinity. The affinity for the AP-1 site by BCD is very much reduced.

## Discussion

It is interesting that all plant bZIP proteins identified to date displayed related but at least three distinct DNA-binding specificity (Izawa *et al.*, 1993). Even though a large number of plant bZIP proteins have been isolated, the mechanisms determining DNA binding specificity are not understood. We used the typical plant G-box binding protein EmBP-1 and C-box-binding protein TGA1a to investigate how the amino acids in the basic-hinge region determine the DNA-binding specificity of plant bZIP proteins. Our study indicates that not only the core basic region but also the hinge region are important in determining plant DNA-binding specificity and that changes of these two regions have an additive effect.

N-terminal basic region or AB region (−31 to −24) is integral part of the basic region as demonstrated by deletion and amino acid substitution analysis of the DNA binding domain (Schindler *et al.*,

1992a,b; Guiltinan and Miller, 1994). However, our gel shift assays indicate that the N-terminal basic region does not play a strong role in determining DNA-binding specificity. This region may contribute to DNA-binding specificity in combination with other regions. For example, the D mutant displayed weaker C-box binding specificity than the ABD mutant. It is also possible that mutations in the N-terminal region of plant bZIP proteins affects stability of DNA-protein complexes in a way analogous to that of nuclear factor-interleukin-6 (NF-IL6) (Brasier and Kumar, 1994).

In GCN4 X-ray crystal structures, 5 base-specific contacts and all but one phosphate backbone contacts are made by the core basic region ($-22$ to $-7$). Of these amino acids, only five amino acid differences are seen between plant G-box binding protein and C-box binding protein. Substitution of the five amino acids in this region changed the DNA binding activity and specificity. The crystal structure of GCN4 complexed with the CRE site (TGACGTCA) has been resolved (König and Richmond, 1993; Keller *et al.*, 1995). The CRE site is identical to the C-box-binding site of plant bZIP proteins. Therefore, DNA-protein contacts seen in the GCN4-CRE site complex are likely applicable to the TGA1a-C-box complex. However, GCN4 binds to AP-1 better than the CRE site while TGA1a binds to C-box (CRE site) better than the AP-1 site, therefore adjustments in GCN4-CRE site structure have to be made to account for protein-DNA interactions of plant bZIP proteins. The CRE and AP-1 sites differ in half-site spacing (Figure 2D). Mutagenesis studies of GCN4 have revealed that the hinge region and some adjacent amino acids in the basic region determine the half-site spacing preference of GCN4 (Johnson, 1993; Kim *et al.*, 1993; Kim and Struhl, 1995). One explanation for this is that the basic regions of these GCN4 mutants may bifurcate at a wider angle from the dimerization domain to accommodate the CRE site which has two central base pairs for half-site spacing (König and Richmond, 1993; Keller *et al.*, 1995).

Sequence alignment indicated that the hinge region of plant C-box-binding proteins are highly conserved (Figure 1). This implies that this region also has an important function. Our mutagenesis results indicate that this region is important in determining DNA-binding specificity of plant bZIP proteins.

Studies of GCN4 and GCN4-C/EBP chimeras have demonstrated that the hinge region can change half-site spacing (Johnson, 1993; Kim *et al.*, 1993; Kim and Struhl, 1995). The binding specificity of two bZIP proteins, chicken VBP and *Drosophila* Giant, are also influenced by the hinge region at positions $\pm 4$ of DNA binding site (we denote the central two base pairs as $\pm 0$) (Haas *et al.*, 1995). These results have illustrated the roles of the hinge region in determining DNA-binding specificity to some extent. However, plant G-box and C-box-binding proteins have the same half-site spacing in their binding sites (Figure 2D). Our results with plant bZIP proteins reveal a large effect of the hinge region on DNA-binding specificity. DNA binding sites at positions $\pm 2$ and $\pm 3$ have been changed, for example from G-box (CCACGTGG) to C-box (TGACGTCA). The change possibly also affects DNA-binding sites at position $\pm 4$; however, changes at positions $\pm 4$ can be tolerated by plant bZIP proteins.

It is interesting that, based on the GCN4 crystal structures, the hinge region does not make direct base-specific contact with DNA. The effect of the change in the hinge region must be transmitted to the basic region to display changed binding specificity. One hypothesis to account for this invokes that a change of the hinge region results in a small conformational shift of the entire basic region both in position and in orientation (Kim *et al.*, 1993); however, a detailed proof of the mechanism awaits further structural analysis.

Based on the known bZIP structures (Ellenberger *et al.*, 1992; König and Richmond, 1993; Glover and Harrison, 1995; Keller *et al.*, 1995), among five amino acids that make base-specific contacts with DNA, four are the same in plant bZIP proteins. Only the amino acid at $-15$ differs in plant G-box and C-box-binding proteins. However, a change of the amino acid at $-15$ does not have a large effect on DNA-binding specificity of EmBP-1. Changing four additional amino acids in the core basic region did alter the DNA-binding specificity. These four amino acids are less critical than $-15$ based on X-ray structures and previous mutagenesis data; however, our results suggest that these five amino acids in the core basic region collectively contribute to G-box or C-box DNA-binding specificity.

Our results also indicate that both the core basic and the hinge regions contribute to EmBP-1-binding specificity. When the center or hinge region of EmBP-1 is replaced, G-box-binding activity is decreased. C-box-binding specificity of EmBP-1 mutants with either core basic region or the hinge region of TGA1a is not as strong as mutants with both regions. This result also suggests that combinatorial contributions are very important in determining highly specific DNA binding.

Binding-site selections indicate that none of the mutants exhibited new binding specificity other than to the G-box and C-box. This could indicate that amino acids in EmBP-1 and TGA1a have been evolutionarily selected to be compatible only with G-box or C-box-binding specificity. Amino acid sequence alignment indicates that G-box and C-box binding proteins have their own highly conserved amino acids (boxed residues in Figure 1); this also is consistent with the notion that these G-box- or C-box-specific amino acids in plant bZIP DNA-binding domains may collectively contribute to G-box or C-box-binding specificity.

Even though crystal structures have been resolved for several bZIP proteins (Ellenberger *et al.*, 1992; König and Richmond, 1993; Glover and Harrison, 1995), mutagenesis studies have revealed a complexity of DNA-protein interactions of bZIP proteins not easily predicted from the structures. Previously, bZIP proteins were thought to be a simple system for the study of protein-DNA recognition because the region that interacts with DNA depends only on an $\alpha$-helix, which is not packed closely with other parts of the protein (König and Richmond, 1993). However, the basic regions of bZIP proteins appear to be in an unfolded state and only assume highly structured $\alpha$-helices upon binding to specific high-affinity sites (O'Neil *et al.*, 1990). Mutagenesis and molecular modeling of GCN4 suggest that adaptability at the protein-DNA interface is an important aspect of sequence recognition by bZIP proteins (Kim *et al.*, 1993). Our results show that longer-range effects of hinge region amino acids act together with the amino acids of core basic region to determine the final binding preference of this class of DNA-binding proteins.

## Acknowledgements

We thank David Lederer, Shannon Bailey, and Trisha Hughes for excellent assistance. This work was supported by grants to M.J.G. from the National Science Foundation (MCB 920 6095) and from The Pennsylvania State University, College of Agricultural Sciences (CRIS 3278).

## References

Brasier, A.R. and Kumar, A. 1994. Identification of a novel determinant for basic domain-leucine zipper DNA binding activity in the acute-phase inducible nuclear factor-interleukin-6 transcription factor. J. Biol. Chem. 269: 10341–10351.

de Pater, S., Katagiri, F., Kijne, J. and Chua, N.-H. 1994. bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter. Plant J. 6: 133–140.

Ellenberger, T.E., Brandel, C.J., Struhl, K., Harrison, S.C. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted a helices: crystal structure of the protein-DNA complex. Cell 71: 1223–1237.

Foster, R., Izawa, T. and Chua, N.-H. 1994; Plant bZIP proteins gather at ACGT elements. FASEB J. 8: 192–200.

Glover, J.N. and Harrison, S.C. 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. Nature 373: 257–261.

Guiltinan, M.J. and Miller, L. 1994. Molecular characterization of the DNA binding and dimerization domains of the bZIP transcription factor, EmBP-1. Plant Mol. Biol. 26: 1041–1053.

Guiltinan, M.J., Marcotte, W.R.J. and Quatrano, R.S. 1990. A plant leucine zipper protein that recognizes an abscisic acid response element. Science 250: 267–271.

Haas, N.B., Cantwell, C.A., Johnson, P.F. and Burch, J.B.E. 1995. DNA-binding specificity of the PAR basic leucine zipper protein VBP partially overlaps those of the C/EBP and CREB/ATF families and is influenced by domains that flank the core basic region. Mol. Cell. Biol. 15: 1923–1932.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77: 51–59.

Izawa, T., Foster, R., Chua, N.-H. 1993. Plant bZIP protein DNA binding specificity. J. Mol. Biol. 230: 1131–1144.

Izawa, T., Foster, R., Nakajima, M., Shimamoto, K. and Chua, N.H. 1994. The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. Plant Cell 6: 1277–1287.

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.

Katagiri, F., Lam, E. and Chua, N.-H. 1989. Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature 340: 727–730.

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.

Kim, J. and Struhl, K. 1995. Determinants of half-site spacing preferences that distinguish AP-1 and ATF/CREB bZIP domains. Nucl. Acids Res. 23: 2531–2537.

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. USA 90: 4513–4517.

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.

Menkens, A.E., Schindler, U. and Cashmore, A.R. 1995. The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins. Trends Biochem. Sci. 20: 506–510.

Niu, X. and Guiltinan, M.J. 1994. DNA binding specificity of the wheat bZIP protein EmBP-1. Nucl. Acids Res. 22: 4969–4978.

Niu, X., Adams, C.C., Workman, J.L. and Guiltinan, M.J. 1996: Binding of the wheat basic leucine zipper protein EmBP-1 to nucleosomal binding sites is modulated by nucleosome positioning. Plant Cell 8: 1569–1587.

O'Neil, K.T., Hoess, R.H. and Degrado, W.F. 1990. Design of DNA-binding peptides based on the leucine zipper motif. Science 249: 774–778.

Oyama, T., Shimura, Y. and Okada, K. 1997. The *Arabidopsis* HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. Genes Devel. 11: 2983–2995.

Schindler, U., Beckmann, H. and Cashmore, A.R. 1992a. TGA1 and G-box binding factors: two distinct classes of *Arabidopsis* leucine zipper proteins compete for the G-box-like element TGACGTGG. Plant Cell 4: 1309–1319.

Schindler, U., Terzaghi, W., Beckmann, H., Kadesch, T. and Cashmore, A.R. 1992b. DNA binding site preferences and transcriptional activation properties of the Arabidopsis transcription factor GBF1. EMBO J. 11: 1275–1289.

Schmidt, R.J., Burr, F.A., Aukerman, M.J. and Burr, B. 1990. Maize regulatory gene opaque-2 encodes a protein with a 'leuzine-zipper' motif that binds to zein DNA. Proc. Natl. Acad. Sci. USA 87: 46–50.

Schmidt, R.J., Burr, F.A. and Burr, B. 1987. Transposon tagging and molecular analysis of the maize regulatory locus opaque-2. Science 238: 960–963.

Suckow, M., von Wilcken-Bergmann, B. and Müller-Hill, B. 1993a. The DNA binding specificity of the basic region of the yeast transcriptional activator GCN4 can be changed by substitution of a single amino acid. Nucl. Acids Res. 21: 2081–2086.

Suckow, M., von Wilcken-Bergmann, B. and Müller-Hill, B. 1993b. Identification of three residues in the basic regions of the BZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding. EMBO J. 12: 1193–1200.

Suckow, M., Madan, A., Kisters-Woike, B., von Wilcken-Bergmann, B. and Müller-Hill, B. 1994a. Creating new DNA binding specificities in the yeast transcription factor GCN4 by combining selected amino acid substitutions. Nucl. Acids Res. 22: 2198–2208.

Suckow, M., Schwamborn, K., Kisters-Woike, B., von Wilcken-Bergmann, B. and Müller-Hill, B. 1994b. Replacement of invariant bZIP residues with the basic region of the yeast transcriptional activator GCN4 can change its DNA binding specificity. Nucl. Acids Res. 22: 4395–4404.

Suckow, M., Lopata, M., Seydel, A., Kisters-Woike, B., von Wilcken-Bergmann, B. and Müller-Hill, B. 1996. Mutant bZIP-DNA complexes with four quasi-identical protein-DNA interfaces. EMBO J. 15: 598–606.

Tabata, T., Takase, H., Takayama, S., Mikami, K., Nakatsuka, A., Kawata, T., Nakayama, T. and Iwabuchi, M. 1989. A protein that binds to a cis-acting element of wheat histone genes has a leucine zipper motif. Science 245: 965–967.

Tabata, T., Nakayama, T., Mikami, K. and Iwabuchi M. 1991. HBP-1a and HBP-1b: leucine zipper-type transcription factors of wheat. EMBO J. 10: 1459–1467.

Walsh, J., Waters, C.A. and Freeling, M. 1998. The maize gene liguleless2 encodes a basic leucine zipper protein involved in the establishment of the leaf blade-sheath boundary. Genes Devel. 12: 208–218.