

# DNA binding specificity of the wheat bZIP protein EmBP-1

Xiping Niu and Mark J. Guiltinan\*

519 Wartik Laboratory, Department of Horticulture, The Biotechnology Institute, The Center for Gene Regulation, Intercollege Graduate Program in Plant Physiology, The Pennsylvania State University, University Park, PA 16802, USA

Received August 12, 1994; Revised and Accepted October 19, 1994 GenBank accession nos: U07933, M62893 and U10466

## ABSTRACT

**EmBP-1 is a wheat DNA binding protein of the basic leucine zipper (bZIP) class of transcription factors implicated in the mechanisms of abscisic acid mediated gene activation. Understanding the role of EmBP-1 in regulating gene transcription requires elucidation of its DNA binding specificity. The binding of EmBP-1 was studied using gel shift selection of DNA from random sequence pools. DNA binding sites were identified by sequencing of a selected pool and by cloning and sequencing individual sites. The binding sites were compared by mobility shift assay and DNase I footprinting, which show that EmBP-1 binds to a family of sequences with varying degrees of affinity. The highest affinity site bound by EmBP-1 is the palindrome GCCACGTGGC. EmBP-1 also binds several other sequences with high affinity, however most of these are asymmetric. While nearly all sequences bound by EmBP-1 contain an ACGT core sequence, EmBP-1 can also bind at least two sites with altered cores. These results provide a basis for comparing the DNA binding specificity of EmBP-1 with those of other plant bZIP proteins and provide insight into the possible target sites which EmBP-1 might bind *in vivo*.**

## INTRODUCTION

The basic-leucine zipper (bZIP) proteins are a major class of eukaryotic transcription factors which have been identified in organisms ranging from yeast to plants and mammals [1]. The bZIP proteins contain conserved DNA binding and dimerization domains in which the basic region contacts the DNA recognition site and the leucine zipper region forms an efficient dimerization interface [2,3]. Mutational analysis [4–6], domain swapping [7,8] and cross linking via disulfide bonds [9] have demonstrated that the DNA binding specificity of bZIP proteins is primarily determined by the amino acid sequence of the basic region. However, the hinge and leucine zipper regions as well as residues outside the basic and leucine zipper domain may also be involved in determining binding specificity by juxtaposition of the basic regions or other mechanisms [8,10–13]. Thus, the sequence of

the basic region alone is not sufficient for the prediction of DNA binding specificity of this class of proteins.

To date, over 22 plant bZIP proteins have been isolated [14]. Amino acid sequence comparisons show that the basic regions of plant bZIP proteins show more homology to each other than to those of other organisms, indicating that they represent an evolutionarily related subfamily. Consistent with this conservation, all plant bZIP proteins can bind sequences with a conserved ACGT core even though they may do so with different affinities. When the bases flanking the core ACGT are examined for their importance in protein interaction, the binding specificities of the plant bZIP proteins appear quite diverse [15–17]. A recent study of 10 plant bZIP proteins directly compared DNA binding specificities under uniform conditions and led to the identification of three distinct groups based on their binding sequence preferences [18]. While these results are informative, they are limited by the number of oligonucleotides of defined sequences which were tested, for example asymmetric and non-ACGT core DNA sequences were not examined extensively.

The wheat bZIP protein EmBP-1 was isolated by screening a cDNA expression library with a DNA probe containing the abscisic acid response element or ABRE, however EmBP-1 can also bind to sequences other than the ABRE [19,20]. DNA-binding proteins isolated by this method may bind the DNA probe used in a screen *in vitro* but have different functional targets *in vivo*. It is therefore essential to characterize the DNA binding specificity of EmBP-1 to determine if the ABRE is a high affinity site and thus a likely target site *in vivo*. Additionally, if DNA binding specificity is accurately defined for a number of related family members, comparison of the differences between their DNA binding domains may provide knowledge of the general relationships governing basic region–DNA interactions.

Previous studies [18,19] showed that EmBP-1 binds Em1a (GACACGTGGC), Hex (GTGACGTGGC) and four palindromic sequences ( $-4G/T$   $-3C/A$   $-2C$   $-1A$   $-0C$   $+0G$   $+1T$   $+2G$   $+3G/T$   $+4C/A$ ) but has very low affinity for the palindromic sequences that do not include C and G at positions  $-2$  and  $+2$  respectively (nomenclature used in this paper follows that used by Oliphant [21] and Izawa [18]). Dissociation constants of 4

\*To whom correspondence should be addressed

nM for EmBP-1 binding to the G-box (GCCACGTGGC) and C-box (GTGACGTCAC) were determined, making it the most specific G-box binding protein tested [18]. However, because only a limited set of sequences were tested, the extended range of DNA binding specificity of EmBP-1 is unknown. Additionally, the relative affinities of EmBP-1 for its various binding sites have not been determined, nor have the importance of bases outside positions  $-4$  or  $+4$  on the binding of EmBP-1 been assessed.

In order to address these questions, the sequence specificity of EmBP-1 was studied in detail. Selection of high affinity binding sites from random oligonucleotide pools by electrophoretic mobility shift assay (EMSA) was followed by sequence analysis of the selected sites. Quantification of EMSA and DNase I footprinting were used to further define the relative binding affinities of the selected sequences. Our results indicate that while EmBP-1 binds the symmetric sequence GCCACGTGGC with the highest affinity it can also interact strongly with a number of asymmetric sites including the ABRE. While nearly all sequences bound by EmBP-1 contain an ACGT core, EmBP-1 can also bind at least two sites with altered core sequences. This information will be important in understanding the role of EmBP-1 in regulating plant gene expression.

## MATERIALS AND METHODS

### Isolation of EmBP-1 cDNAs

A  $\lambda$ ZAPII cDNA library was constructed using 10  $\mu$ g of poly(A)<sup>+</sup> RNA isolated from immature (stage III) ABA treated wheat embryos and oligo-dT primers as described [22]. cDNA was ligated into *Eco*RI cut  $\lambda$ ZAPII vector DNA (Stratagene, La Jolla, CA) and packaged in GigaPack Gold extracts (Stratagene, La Jolla, CA). By screening this library by hybridization with the original EmBP-1 cDNA GC19 [19], a plasmid (pGCF13) containing a 1.4 kb cDNA insert was isolated (Figure 1B), however based on the EmBP-1 mRNA size as judged from northern blot experiments (1.6–1.7 kb) [19] it was apparently not full length. In order to obtain overlapping and/or full length cDNA isolates, fragments from pGCF13 were used to screen  $1.2 \times 10^6$  phage from a  $\lambda$ gt11 cDNA library [19] by hybridization at high stringency [23]. 15 partial length cDNA clones were isolated and two of these (pGCNN5.1 and pGCQ7.1) were shown to extend past the 5' end of the pGCF13 cDNA. An additional clone (pGCN46) was isolated which shows 100% identity with pGCNN5.1 over 514 bp of overlapping sequence and which contains an intact 3' end including a poly A tract. By assembly of pGCNN5.1 and pGCN46, we have constructed a plasmid containing the full length sequence designated pGCNN546.

### Expression and purification of recombinant EmBP-1

To construct a plasmid containing the full length EmBP-1a cDNA sequence as presented in Figure 1B, *Eco*RI–*Bam*HI DNA fragments from two overlapping clones (pGCNN5.1 and pGCN46) were ligated at their single internal *Bam*HI restriction sites (Figure 1A) and into the *Eco*RI site of pGEM7Zf (Promega, Madison, WI) creating plasmid pGCNN546. For the purpose of expression in *E. coli*, a second similar fusion was made which contains less 3' untranslated DNA and no poly A tract by fusing DNA from two overlapping partial cDNA clones pGCNN5.1 (from 1 to 1145) and pGCN16 (from 635 to 1291) at their *Bam*HI sites resulting in plasmid pGCN516 (Figure 1A). The insert DNA

from pGCN516 was cut with *Sac*I, partially digested with *Nco*I, Klenow repaired and inserted in frame into a histidine fusion expression vector pV2b [24] at a Klenow filled *Eco*RI site resulting in the expression vector pXN11 which was transformed into JM109. This resulted in the in frame fusion of sequences encoding the amino acids MTMITPSSHHHHHLEDRSPGTELEF to the EmBP-1a amino acid sequence which we have designated His6-EmBP-1a (Figure 1C). For induction of His6-EmBP-1a protein expression, cells were grown in LB to an O.D.<sub>600</sub> of about 0.4–0.5 and induced with 0.4mM isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG). Cells were grown for another 1–3 h and harvested by centrifugation. The histidine fusion protein was extracted and purified over a nickel ion column according to the manufacturer's instructions (Novagen, Madison, WI). The eluate containing His6-EmBP-1a was dialyzed against  $1 \times$ TEB (20 mM Tris–HCl pH 7.5, 1 mM EDTA, 1 mM  $\beta$ -mecaptoethanol and 10% glycerol) and loaded on a heparin–agarose column pre-equilibrated with  $1 \times$ TEB + 0.1 M NaCl. The proteins were then eluted with a gradient of 0.1–1 M NaCl in  $1 \times$ TEB.

Proteins were separated by SDS–PAGE and electrophoretically transferred to nitrocellulose filters in transfer buffer (3.03 g Tris, 14.4 g glycine and 200 ml methanol per liter) using Bio-Rad Mini Trans-Blot cell according to manufacturer's instruction (Bio-Rad, Richmond, CA). Filters were blocked with 5% nonfat dry milk in  $1 \times$ binding buffer (25 mM Hepes pH 7.9, 25 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.5 mM DTT), then incubated with <sup>32</sup>P labeled ABRE [19] for 16 h at 4°C in  $1 \times$ binding buffer containing 50  $\mu$ g/ml denatured salmon sperm DNA and washed four times in  $1 \times$ binding buffer for 7.5 min each wash at 4°C. Filters were imaged by autoradiography at  $-80^\circ\text{C}$  with intensifying screens and Kodak XAR5 film.

### EMSA and quantification

Performed as previously described [25].

### SAAB selection

Single stranded synthetic oligonucleotides containing random sequences (Table 1) were first made double stranded and labeled at 37°C for 30 min in a 20  $\mu$ l reaction containing 25 ng of oligo, 224 ng of the pUC reverse primer, 25 mM Tris–HCl pH 8.0, 5 mM MgCl<sub>2</sub>, 2 mM DTT, 200 mM Hepes pH 6.6, 25  $\mu$ M of dATP, dTTP and dGTP each, 50  $\mu$ Ci dCTP at 6000 Ci/mmol, and 5 units of Klenow. DNA was purified on 12% polyacrylamide gel and electroeluted in  $0.1 \times$ TBE [23]. Specific activities were in the range of  $8.4 \times 10^8$  c.p.m./ $\mu$ g DNA. About 100,000 c.p.m. of DNA was used in the 20  $\mu$ l binding reaction which contained 40 ng purified His6-EmBP-1a protein, 1  $\mu$ g dI:dC, 0.25% TritonX-100 and 8  $\mu$ g BSA in  $1 \times$ binding buffer IV [25] or with KCl at concentrations indicated in the text. After separation on EMSA gels the bound DNA was excised and electroeluted into dialysis tubing in  $0.1 \times$ TBE. 20–50 c.p.m. was amplified by PCR with PUC forward and reverse primers in 100  $\mu$ l of reaction buffer for 35 cycles (30 sec at 93°C, 30 sec at 45°C, 45 sec at 70°C). For the further rounds of selection, DNA was labeled by *Taq* DNA polymerase at 70°C and gel purified as described above.

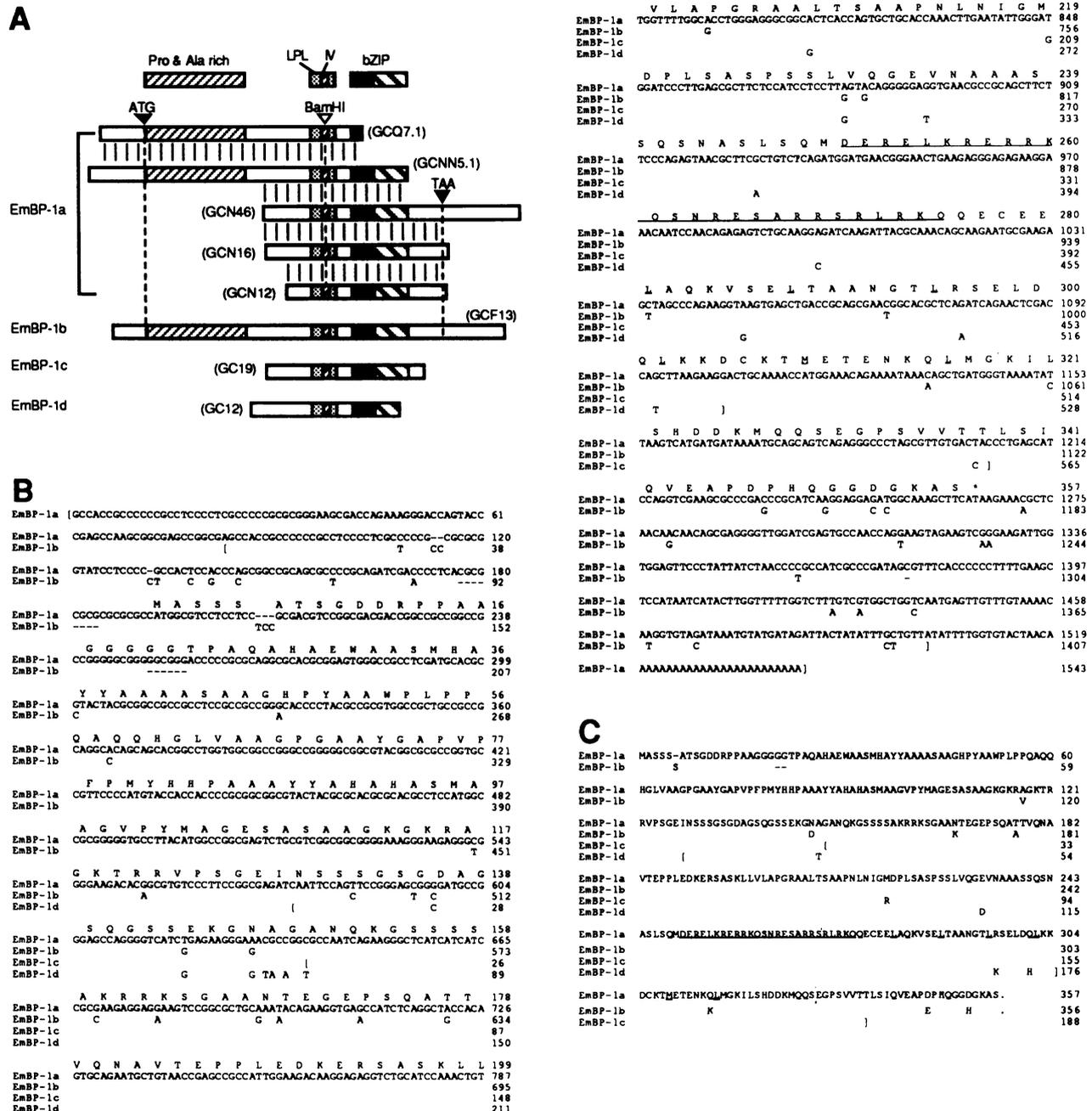
To sequence the PCR products, forward or reverse primers were labeled by T4 polynucleotide kinase (PNK) in the reaction as recommended by the manufacturer (Promega, Madison, WI), the specific activity of end labeled primers were in the range of  $10^9$  c.p.m./ $\mu$ g. The sequencing reactions were essentially as

previously described [26]. Selected sequences from oligo 2 after five rounds were digested with *Xba*I and *Hind*III and cloned into similarly digested PUC18 vector.

**DNase I footprinting**

The probes for DNase I footprinting were prepared by PCR amplification of cloned DNA binding sites with PUC forward

and reverse primers. For top strand footprinting, PUC forward sequencing primer was end-labeled with PNK kinase. For the bottom strand, PUC reverse primer was end-labeled. The conditions for PCR and end-labeling reaction were the same as above except that PCR reactions were scaled to 10  $\mu$ l. The probes were purified on 8% polyacrylamide gels and electroeluted into dialysis tubing in 0.1  $\times$  TBE. The probe concentrations were about



**Figure 1.** cDNA sequences encoding EmBP-1s and deduced amino acid sequences. (A) Diagrams of EmBP-1 family cDNAs. Multiple isolates of EmBP-1a shown share 100% sequence identity over their overlapping regions (indicated with vertical lines). Start codons (ATG), basic and leucine zipper domains, stop codons, a conserved *Bam*HI restriction site, the proline-alanine rich domain, the unique proline-interspersed leucine repeat (LPL) and the conserved motif IV [33] are indicated. Clone numbers are indicated in parenthesis. (B) cDNA sequences of the EmBP-1 family. Nucleotide numbering is from the first nucleotide of each sequence. The bases differing from EmBP-1a are shown. - indicates deletion of a nucleotide, [ and ] indicate the start and end of each sequence. The deduced amino acid sequence of EmBP-1a is shown on top, numbered from the first amino acid. The basic domain and leucine residues are underlined (methionine replaces one of the leucines). GenBank accession numbers of EmBP-1a, b, c and d are U07933, M62893, M63999 and U10466 respectively. (C) Amino acid sequences of the EmBP-1 family. Numbering is from the first amino acids, labeling is as in (B).

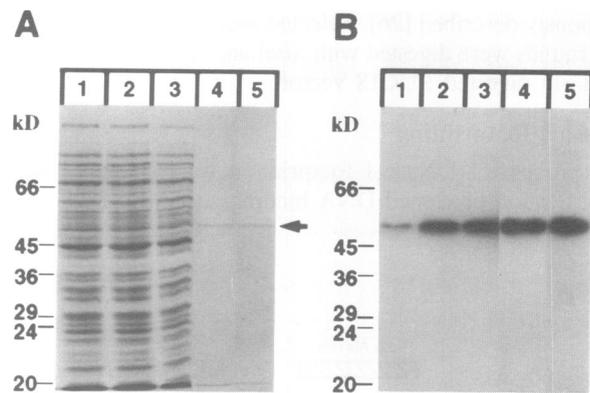
1,000–5,000 c.p.m./ $\mu$ l eluate with a specific activity of  $1.3 \times 10^8$  c.p.m./ $\mu$ g. 1,500–2,000 c.p.m. of probes were incubated with His6-EmBP-1a at room temperature for 20 min in 200  $\mu$ l of  $1 \times$  binding buffer IV, 0.1 mg/ml BSA and 2  $\mu$ g/ml native salmon sperm DNA. DNase I digestion was performed for 2 min with addition of 5  $\mu$ l of DNase I at 0.1 mg/ml. The reactions were stopped with 700  $\mu$ l ice cold stop solution (mixture of 650  $\mu$ l ethanol, 50  $\mu$ l saturated  $\text{NH}_4\text{Ac}$  and 5  $\mu$ g of yeast tRNA) and returned to ice immediately. DNA was pelleted by centrifugation and washed with 70% ethanol. DNA was then resuspended in 5  $\mu$ l sequencing gel loading buffer and 3  $\mu$ l was loaded on sequencing gels.

## RESULTS

### Isolation of full length cDNA clones encoding EmBP-1

The original EmBP-1 partial cDNAs were obtained from a  $\lambda$ gt11 expression library screen which resulted in the isolation of two clones ( $\lambda$ GC19 and  $\lambda$ GC12) encoding nearly identical bZIP proteins which bind to the abscisic acid response element [19]. In order to isolate full length clones encoding EmBP-1, additional screens were performed on two different cDNA libraries (see Materials and Methods). Schematic drawings of the pertinent clones from these screens are presented in Figure 1A. For simplicity, the cDNA sequences from clones pGCNN546, pGCF13,  $\lambda$ GC19 and  $\lambda$ GC12 as well as the proteins they encode have been designated EmBP-1a, b, c and d respectively, and these will be referred to collectively as EmBP-1 cDNAs and proteins. As shown in Figure 1B, within the EmBP-1a and EmBP-1b cDNA sequences, single long coding regions exist which are in frame with those of the original EmBP-1c and EmBP-1d cDNA sequences. While some DNA sequence divergence is observed between the various isolates, the encoded amino acid sequences show only a few differences (Figure 1C). The isolation of 4 classes of cDNAs encoding EmBP-1 is consistent with the fact that upon Southern genomic analysis, multiple bands hybridize with a EmBP-1 probe, allowing the mapping of 7 different genomic loci in the wheat genome [27].

An examination of the amino acid sequences of EmBP-1 family members demonstrates strong amino acid sequence conservation except at a few positions (Figure 1C). The basic regions among these family members are identical. The percent amino acid conservation among these EmBP-1 family members ranges from 98.9% to 96.6% which is as high or higher than seen between other bZIP family members. The N-terminal third of EmBP-1 is particularly rich in proline and alanine residues (58% of total Pro+Ala from residues 1–119), characteristics of transcriptional activation domains found in a number of other DNA binding proteins [28–30]. Present in the central third of the protein is a unique leucine heptad repeat element which contains interspersed proline residues in each of four repeat elements. Present within this repeat domain is a heptad motif (NLNIGND, aa residues 214–220) which is conserved between wheat EmBP-1, HBP-1a, Arabidopsis GBF-1, and tomato GBF-4, GBF-9 and GBF-12 [31–33]. The functional significance of this domain is not known at this time, although it is not required for DNA binding or dimerization [25]. A second conserved amino acid block is partially conserved between EmBP-1, Arabidopsis GBF-1 and tomato GBF-4 and GBF-12 (SGSG--G, aa residues 132–138). The potential role of these conserved blocks in protein–protein interactions has previously been discussed [33].



**Figure 2.** Expression and purification of His6-EmBP-1a. (A) 10% SDS–polyacrylamide gel. Lane 1, crude protein extract from non-induced *E. coli* culture. Lanes 2 and 3, protein extract after induction with IPTG for 1 and 3 h. Cell concentrations were adjusted so that extracts from approximately equal numbers of cells were loaded. Lane 4, the eluate after nickel ion chelate chromatography. Lane 5, the concentrated pool after heparin–agarose chromatography. Sizes of molecular weight marker are indicated on the left, the position of His6-EmBP-1a is indicated with an arrow. (B) Southwestern blot of an identical gel as in (A) probed with a radiolabeled DNA fragment containing the ABRE target sequence for EmBP-1 [19].

The C-terminal third of the protein contains the bZIP domains as previously described [19].

### Overexpression of recombinant EmBP-1

In order to obtain purified recombinant EmBP-1 protein for DNA binding studies, we overexpressed EmBP-1a with an N-terminal 6 residue histidine fusion in *E. coli*. As shown in Figure 2, cells expressing His6-EmBP-1a did not contain large amounts of the fusion protein (lanes 1–3), however, the His6-EmBP-1a protein was readily purified under non-denaturing conditions by nickel ion chelate chromatography. At stringent wash conditions (100 mM imidazole), only minor contaminants remained (Figure 2A, lane 4). After a second heparin–agarose column, the His6-EmBP-1a protein was nearly pure as evidenced by a single band on a coomassie blue stained SDS gel (Figure 2A, lane 5). The identity and DNA binding activity of His6-EmBP-1a was demonstrated by Southwestern blot analysis using the ABRE probe (Figure 2B). The recombinant His6-EmBP-1a protein retained strong DNA binding affinity for the ABRE probe, even after being denatured with SDS and immobilized on nitrocellulose filters (Figure 2B). The apparent molecular weight of His6-EmBP-1a as judged by relative mobility on SDS gels is slightly higher than the predicted weight based on the deduced amino acid sequence (39.8-kD), a consistent feature of EmBP-1 likely to be caused by the strong basic charge of the DNA binding domain. Competitive gel shift assays indicate that the DNA binding activity of the fusion protein is indistinguishable from full-length EmBP-1a translated *in vitro* or expressed in *E. coli* (data not shown).

### Determination of EmBP-1 binding specificity

In order to determine the range of sequences efficiently bound by EmBP-1, random oligonucleotide pools were used as probes for the selection of binding sites by the Selected And Amplified Binding site method (SAAB) [26]. In this method, bound oligonucleotides are fractionated from unbound sequences on

EMSA gels, excised and amplified by PCR. After several rounds of selection, the purified sequences can either be sequenced as a pool to produce a 'SAAB imprint' or they can be cloned and sequenced individually. Three important parameters were controlled when SAAB selection was carried out. First, in order to increase specificity and reduce the number of cycles required for efficient selection, DNA binding was first tested at different salt conditions and then routinely performed at a high salt (e.g. 1 M KCl) which increased binding specificity (see below). Second, unequal coupling efficiencies during DNA synthesis could result in a biased representation in the pool. Third, PCR reactions might differentially amplify sequences in the random pool causing artifacts on SAAB imprinting. For the second and third considerations, control samples from the non-selected random sequence pools were amplified with the same amounts of input DNA as were the bound fractions. Any bias in coupling efficiency or amplification should become apparent in the non-selected controls.

To determine the DNA recognition sequences of EmBP-1 by SAAB, two pools of degenerate oligonucleotides (oligos) were synthesized as shown in Table 1: oligo 1 which has the ACGT core sequence flanked on both sides with 5 random base pairs and oligo 2, which has a continuous stretch of 14 random base pairs. Both oligos are flanked by cloning and PCR priming sites. Binding reactions with oligo 1 were performed at 35 mM, 600 mM and 1 M KCl, and aliquots of bound probes were PCR amplified and retained for sequencing after each of five rounds of selection. Representative autoradiograms depicting DNA sequencing reactions of the SAAB selected pools are presented in Figure 3. Strong selection was clearly apparent at positions 2, 3 and 4 and weaker selection was observed out to position 6 in both 5' and 3' directions (most apparent in Figure 3C). At all the salt conditions, C was selected for the position -2 and G for +2. Only minor differences in selection at positions 3 and 4 were observed between samples bound at the different salt concentrations. However at the high salt condition, SAAB imprints are more clear indicating that the high salt increases the selection specificity (Figure 3C).

The selected sites are not symmetric as evidenced by the differences in base representation at different positions. Detailed examination of SAAB imprinting at high salt reveals that: at position -3, C and A are almost equally selected for on the left half site but on the right half site (position +3) G is strongly selected over T. At position -4, G is selected but is not as strong as C at the corresponding position +4 on the right side. At position -5, G is selected strongly on the left half site but A is selected on the right side (position +5).

Because position +4 on the right half site is selected strongly, we conclude that the EmBP-1 binding site consists of ten base pairs. On the basis of SAAB imprinting at a high salt selection condition, the consensus selected binding site is G(T)<sub>-4</sub> C(A)<sub>-3</sub> C<sub>-2</sub> ACGT G<sub>+2</sub> G(T)<sub>+3</sub> C(A)<sub>+4</sub>. The alternative weakly selected bases are shown in parenthesis. The various combinations

of these selected half sites will generate 10 different binding sites as shown in Table 2. Judging from the band intensity on SAAB imprint, it can be predicted that sites such as GCCACGTGGC are strongly selected while others such as TACACGTGTA are less abundant in the selected pool.

In order to examine further the binding specificity of EmBP-1, SAAB selection was performed on Oligo 2 containing a 14 bp random sequence (Table 1) at high salt conditions (1 M KCl). To determine the sequences of the selected binding sites the PCR products were cloned into pUC18 and 61 isolates were sequenced. Based on sequence differences, 19 distinct groups were identified all of which (excepting two) contain the ACGT core (Figure 4). A large number of the selected binding sites from oligo 2 are not observed in the SAAB imprint using oligo 1. In the selected binding sites from oligo 1, CACGTG is strongly selected while in the sites selected from oligo 2, CACGTG and CACGTG are nearly equally represented. From the selection of oligo 2, two non-ACGT binding sites were also isolated (Figure 4, clones 22 and 45). In both of these sequences, it is the central CG base pairs that differ, CG being replaced either by CGC or GC.

Consistent with results from oligo 1, position 4 is shown to affect DNA binding strongly. Out of 61 clones sequenced, 43 selected binding sites are opposed to the left side of oligo 2 with G and C next to the random region used as positions -4 and -3. In contrast, only 7 binding sites are opposed to the right side of oligo 2, having G and A in positions +3 and +4 of their binding sites (Table 1 and Figure 4). The remaining 11 selected binding sites are centered within the 14 bp random sequence region and do not use the flanking DNA as part of their binding sites.

The four half sites that produce symmetric binding sites recognized by EmBP-1 (Table 2, #1, #5, #8 and #10) we have designated 'symmetric half sites' consistent with the findings of Izawa [18]. It is clear that some of the selected sites identified in Figure 4 are composed of one 'symmetric half site' combined with half sites that EmBP-1 can not bind when made symmetric. Half sites which can not be bound by EmBP-1 when made symmetric but which can be bound when combined with a symmetric half site we have designated 'dependent half sites'. According to the concepts of symmetric and dependent half sites, the binding sites selected from oligo 2 are organized in Table 3. Combinatorial associations of four symmetric half sites will generate 10 different DNA binding sites, only 7 of which were cloned from the selected pool of oligo 2 (#1-7 in Table 3). According to the prediction made by Williams [16], one of the strong class I half sites for type A plant bZIP proteins is G<sub>-4</sub> C<sub>-3</sub> C<sub>-2</sub> A<sub>-1</sub> C<sub>-0</sub>. Consistent with this observation, we found that the GCCAC is found in 11 combinations of other half sites (#1-4, #8-14 in Table 3), whereas for the weaker half site T<sub>-4</sub> A<sub>-3</sub> C<sub>-2</sub> A<sub>-1</sub> C<sub>-0</sub>, only 2 combinations with other half sites were selected (#4 and #17 in Table 3). Among the dependent half sites, most include either G<sub>+0</sub> T<sub>+1</sub> C<sub>+2</sub> A<sub>+3</sub> T<sub>+4</sub>, G<sub>+0</sub> T<sub>+1</sub> C<sub>+2</sub> A<sub>+3</sub> C<sub>+4</sub> or G<sub>+0</sub> T<sub>+1</sub> C<sub>+2</sub> A<sub>+3</sub> G<sub>+4</sub>. The other EmBP-1 dependent half site worthy of note is G<sub>+0</sub> T<sub>+1</sub> A<sub>+2</sub> A<sub>+3</sub> G<sub>+4</sub> which exists in naturally occurring *cis*-elements *ocs* and *as-1* [34,35], and is also a strong half site for the yeast bZIP protein GCN4 [36]. The ability of the dependent sequences to generate a specific DNA binding site with other defined sequences may reflect some degree of flexibility associated in bZIP proteins/DNA interactions.

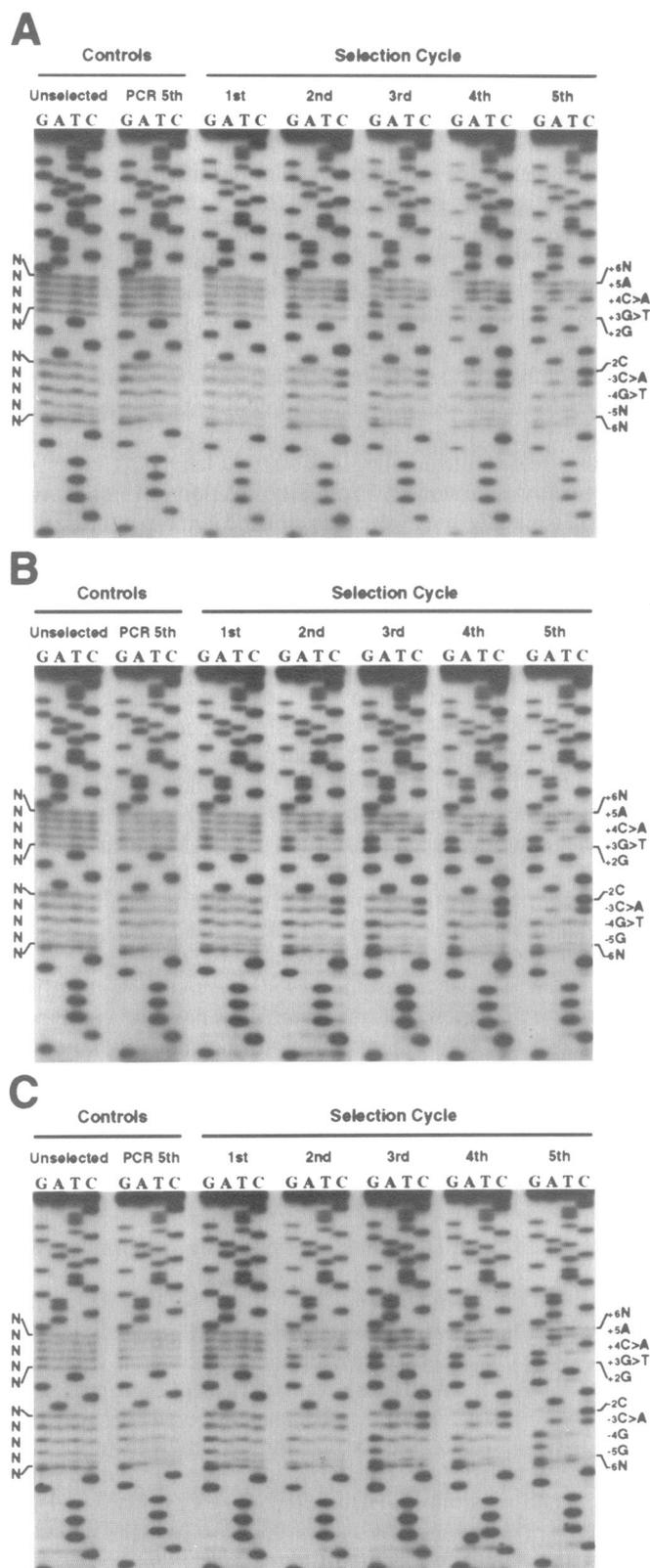
In the two non-ACGT core sites, one (#19) has three central base pairs (CGC) and is flanked by two strong symmetric half

Table 1. Sequences of mixed oligonucleotides used in SAAB selection

#	PUC forward primer	Re L <sup>a</sup>	LF <sup>b</sup>	Random seq.	RF <sup>b</sup>	Re R <sup>a</sup>	PUC Reverse primer
1	GTAAACGACGGCCAGT	AAGCTT	TGC	(N) <sub>5</sub> ACGT(N) <sub>5</sub>	GAC	AAGCTT	GTCATAGCTGTTCTCTG
2	GTAAACGACGGCCAGT	AAGCTT	TGC	(N) <sub>14</sub>	GAC	TCTAGA	GTCATAGCTGTTCTCTG

<sup>a</sup>Restriction enzyme sites Re L is *Hind*III, Re R is *Hind*III for oligo1 and *Xba*I for oligo 2.

<sup>b</sup>LF and RF indicate left and right flanking sequences.



**Figure 3.** SAAB imprinting of EmBP-1 binding specificity. DNA sequence analysis of oligonucleotide pools. The first two sets of lanes contain the unselected oligo and PCR reaction controls (unselected and PCR 5th). Sequences from each of five rounds of selection are presented. Positions of random bases and the consensus selected nucleotides are indicated on the left and right respectively. (A) Binding reactions performed at 35 mM KCl. (B) Binding reactions performed at 600 mM KCl. (C) Binding reactions performed at 1 M KCl.

sites ( $G_{-4} C_{-3} C_{-2} A_{-1}$  on the left and  $T_{+1} G_{+2} G_{+3} C_{+4}$  on the right). The other non-ACGT site (# 18) has two central base pairs (GC) flanked by a strong symmetric half site on the left ( $G_{-4} C_{-3} C_{-2} A_{-1}$ ) and a symmetric half site on the right ( $T_{+1} G_{+2} G_{+3} A_{+4}$ ). These results show that EmBP-1 can tolerate sequence variation in the center of DNA binding sites, a feature shared with other bZIP proteins [35,37,38].

#### Analysis of selected binding sites by EMSA and DNase I footprinting

Based on selection alone, the possibility that some of these sequences might exhibit low affinity and/or non-specific binding by His6-EmBP-1a can not be ruled out. In order to test the relative binding affinity of selected sites individually, EMSA was performed on each of the 19 classes of consensus binding sites (Figure 4). The sites indicated in Figure 4 (\*) representing each consensus class and the pUC18 polylinker control were amplified by PCR with a single end-labeled primer and incubated with equal amounts of protein for comparison of relative binding affinity by EMSA as presented in Figure 5A.

Quantification of the amount of DNA bound by EmBP-1 in three replicate experiments show that His6-EmBP-1a binds the palindromic sequence GCCACGTGGC with the highest affinity (Figure 5B, # 1). His6-EmBP-1a can also bind other sites with affinity comparable to that of the highest affinity site (Figure 5B, # 9 and # 16). The high affinity sites (Figure 5B, # 2–4, # 8, # 10–12, # 15 and # 17) also include naturally occurring sequences found in plant promoters such as the ABRE, G-1A and Hex [19,31,32]. These high affinity sites are either composed of two symmetric half sites (ABRE, # 2 and G-1A, # 3) or a symmetric half site combined with a dependent half site (Hex, # 8). Binding sites that have the dependent half site  $T_{+1} C_{+2} A_{+3}$  or  $T_{+1} A_{+2} A_{+3}$  are generally high affinity sites. One binding site (# 19) without an ACGT core but having three central base pairs CGC flanked by two strong symmetric half sites can be bound with respectable affinity. Among the relatively low affinity binding sites (# 5–7, # 13, # 14, # 18 and # 19), the lowest affinity binding site (# 14) is generated by replacing  $G_{+2}$  in the highest affinity binding site with  $T_{+2}$ . Inconsistent with the previous results [16,18], our quantification of EMSA indicates that the symmetric site GACACGTGTC (# 5) is a lower affinity binding site, which has also been demonstrated in DNase I footprinting (see below) and gel competition assays (not shown).

In order to verify the bases bound by EmBP-1 and to further assess the relative binding affinities of the various sequences, DNase I footprinting was performed on the selected sites. The results of DNase I footprinting shown in Figure 6A and 6B indicate that in all cases, the protected regions center on the ACGT core element (Figure 6C). Judged by relative strength and extent of protection at different protein concentrations, these

**Table 2.** EmBP-1 recognition sites predicted by SAAB imprint analysis. Strongly selected bases are in bold

#	-4-3-2-1-0	+0+1+2+3+4
1	<b>G C C A C</b>	<b>G T G G C</b>
2	<b>G C C A C</b>	<b>G T G T C</b>
3	<b>G C C A C</b>	<b>G T G G A</b>
4	<b>G C C A C</b>	<b>G T G T A</b>
5	<b>G A C A C</b>	<b>G T G T C</b>
6	<b>G A C A C</b>	<b>G T G G A</b>
7	<b>G A C A C</b>	<b>G T G T A</b>
8	<b>T C C A C</b>	<b>G T G G A</b>
9	<b>T C C A C</b>	<b>G T G T A</b>
10	<b>T A C A C</b>	<b>G T G T A</b>

Clone#	-4321001234+	Consensus	
16-1	<b>AAGCTTT</b> GCCACGTGGC	AAGGGC <b>GA</b> CTCTAGA	1
16-2*	<b>AAGCTTT</b> GCCACGTGGC	ATAACC <b>GA</b> CTCTAGA	GCCACGTGGC
21	<b>AAGCTTT</b> GCCACGTGGC	ACCTTC <b>GA</b> CTCTAGA	
28	<b>AAGCTTT</b> GCCACGTGGC	TTCTCT <b>GA</b> CTCTAGA	
47	<b>AAGCTTTGC</b> ACAT	GCCACGTGGC <b>GA</b> CTCTAGA	
8	<b>AAGCTTT</b> GCCACGTGTC	GTTCTG <b>GA</b> CTCTAGA	2
41r	<b>TCTAGAGTC</b> ATGT	GCCACGTGTC <b>GCA</b> AAGCTT	GCCACGTGTC
19*	<b>AAGCTTT</b> GCCACGTGTC	TTCTCT <b>GA</b> CTCTAGA	
25	<b>AAGCTTT</b> GCCACGTGTC	CCATCA <b>GA</b> CTCTAGA	
6	<b>AAGCTTTGC</b>	GCCACGTGTC <b>CTTA</b> <b>GA</b> CTCTAGA	
59	<b>AAGCTTTGC</b>	GCCACGTGTC <b>CTAT</b> <b>GA</b> CTCTAGA	
27	<b>AAGCTTT</b> GCCACGTGGA	ATTGAA <b>GA</b> CTCTAGA	3
57	<b>AAGCTTT</b> GCCACGTGGA	AAACAA <b>GA</b> CTCTAGA	GCCACGTGGA
5	<b>AAGCTTT</b> GCCACGTGGA	TTATAA <b>GA</b> CTCTAGA	
13	<b>AAGCTTT</b> GCCACGTGGA	CTGCAA <b>GA</b> CTCTAGA	
37	<b>AAGCTTT</b> GCCACGTGGA	CACAGA <b>GA</b> CTCTAGA	
44-2	<b>AAGCTTTGC</b> GATCAT	GCCACGTG <b>GA</b> <b>CTCTAG</b>	
50*	<b>AAGCTTTGC</b> ACCTCT	GCCACGTG <b>GA</b> <b>CTCTAGA</b>	
12	<b>AAGCTTTGC</b> ATTAGC	GCCACGTG <b>GA</b> <b>CTCTAGA</b>	
38	<b>AAGCTTT</b> GCCACGTGTA	ATATGT <b>GA</b> CTCTAGA	4
48	<b>AAGCTTT</b> GCCACGTGTA	AAACTG <b>GA</b> CTCTAGA	GCCACGTGTA
2	<b>AAGCTTT</b> GCCACGTGTA	CCATTT <b>GA</b> CTCTAGA	
15	<b>AAGCTTT</b> GCCACGTGTA	CTTATA <b>GA</b> CTCTAGA	
46*	<b>AAGCTTT</b> GCCACGTGTA	CTCATA <b>GA</b> CTCTAGA	
39-1*	<b>AAGCTTTGC</b>	GACACGTGTC <b>CTTT</b> <b>GA</b> CTCTAGA	5
			GACACGTGTC
61	<b>AAGCTTTGC</b> CATCTT	GACACGTG <b>GA</b> <b>CTCTAGA</b>	6
56	<b>AAGCTTTGC</b> AGTGGG	GACACGTG <b>GA</b> <b>CTCTAGA</b>	GACACGTGGA
39-2*	<b>AAGCTTTGC</b>	GACACGTGGA <b>CCAT</b> <b>GA</b> CTCTAGA	
42*	<b>AAGCTTTGC</b> ATTACA	TCCACGTG <b>GA</b> <b>CTCTAGA</b>	7
			TCCACGTGGA
23	<b>AAGCTTT</b> GCCACGTGTC	CTTGCA <b>GA</b> CTCTAGA	8
29*	<b>AAGCTTT</b> GCCACGTGTC	CTTATA <b>GA</b> CTCTAGA	GCCACGTGTC
3	<b>AAGCTTT</b> GCCACGTGTC	CTCCTC <b>GA</b> CTCTAGA	
34	<b>AAGCTTT</b> GCCACGTGTC	CACTTA <b>GA</b> CTCTAGA	
49	<b>AAGCTTT</b> GCCACGTGTC	CACTTT <b>GA</b> CTCTAGA	
32	<b>AAGCTTT</b> GCCACGTGTC	CATACA <b>GA</b> CTCTAGA	
33	<b>AAGCTTT</b> GCCACGTGTC	CCTTTC <b>GA</b> CTCTAGA	
24	<b>AAGCTTT</b> GCCACGTGTC	CCTATA <b>GA</b> CTCTAGA	
9	<b>AAGCTTT</b> GCCACGTGTC	TGTA <b>GA</b> CTCTAGA	
4	<b>AAGCTTT</b> GCCACGTGTC	TACTAC <b>GA</b> CTCTAGA	
35	<b>AAGCTTT</b> GCCACGTGTC	ATACTC <b>GA</b> CTCTAGA	
7	<b>AAGCTTT</b> GCCACGTGTC	CATGTC <b>GA</b> CTCTAGA	9
11	<b>AAGCTTT</b> GCCACGTGTC	CATGAC <b>GA</b> CTCTAGA	GCCACGTGTC
52*	<b>AAGCTTT</b> GCCACGTGTC	CATCAG <b>GA</b> CTCTAGA	
36	<b>AAGCTTT</b> GCCACGTGTC	CACCAA <b>GA</b> CTCTAGA	
54	<b>AAGCTTT</b> GCCACGTGTC	CACAAA <b>GA</b> CTCTAGA	
1	<b>AAGCTTT</b> GCCACGTGTC	CAACCA <b>GA</b> CTCTAGA	
60	<b>AAGCTTT</b> GCCACGTGTC	ATACCAA <b>GA</b> CTCTAGA	
14	<b>AAGCTTT</b> GCCACGTGTC	GCACGT <b>GAT</b> CTAGA	
58*	<b>AAGCTTTGC</b> ACAGT	GCCACGTGTC <b>ACT</b> CTAGA	10
			GCCACGTGTC
10	<b>AAGCTTT</b> GCCACGTGTC	CATCTC <b>GA</b> CTCTAGA	11
62*	<b>AAGCTTT</b> GCCACGTGTC	CATCAT <b>GA</b> CTCTAGA	GCCACGTGTC
31r*	<b>TCTAGAGTC</b> TAAAC	GCCACGTGTC <b>CAA</b> AAGCTT	12
			GCCACGTGTC
51	<b>AAGCTTT</b> GCCACGTGTC	CTTTAC <b>GA</b> CTCTAGA	13
55*	<b>AAGCTTT</b> GCCACGTGTC	TCTTCT <b>GA</b> CTCTAGA	GCCACGTGTC
40*	<b>AAGCTTT</b> GCCACGTGTC	ACTCTG <b>GA</b> CTCTAGA	14
			GCCACGTGTC
30r*	<b>TCTAGAGTC</b> AGATT	GACACGTGTC <b>CAA</b> AAGCTT	15
20r	<b>TCTAGAGTC</b> TTGAT	GACACGTGTC <b>CAA</b> AAGCTT	GACACGTGTC
43r*	<b>TCTAGAG</b>	TCCACGTGTC <b>CAT</b> AGC <b>GCA</b> AAGCTT	16
			TCCACGTGTC
44-1r*	<b>TCTAGAGTC</b> GTTAG	TACACGTGTC <b>CAA</b> AAGCTT	17
			TACACGTGTC
45*	<b>AAGCTTT</b> GCCACGTGTC	CATATA <b>GA</b> CTCTAGA	18
			GCCACGTGTC
22*	<b>AAGCTTT</b> GCCACGTGTC	ACCCC <b>GA</b> CTCTAGA	19
			GCCACGTGTC

results are consistent with the relative DNA binding affinities of the sequences as determined by gel shift analysis.

It is noteworthy that for some of the sequences (#10, #11, #16 and #17), EmBP-1 causes a change in DNA conformation upon binding as evidenced by the presence of DNase I hypersensitive sites. The sites exhibiting DNase I hypersensitivity are associated with a dependent half site of either T<sub>+1</sub> C<sub>+2</sub> A<sub>+3</sub> G<sub>+4</sub> or T<sub>+1</sub> A<sub>+2</sub> A<sub>+3</sub> G<sub>+4</sub>. In all cases seen, the hypersensitive sites are located at position +5 (Figure 6C). EmBP-1 might bind these sequences with a different conformation causing a change in DNA structure.

### DISCUSSION

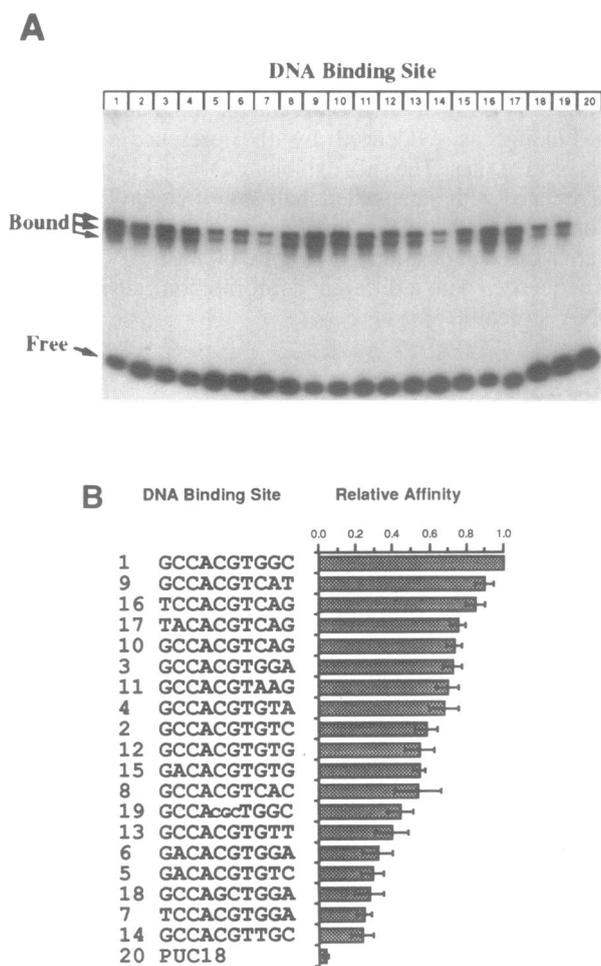
Characterization of the specificity of DNA binding proteins is a necessary prerequisite to the understanding of their functional significance *in vivo*. Previous studies have provided limited information on the binding specificity of EmBP-1 [18,19]. In order to characterize EmBP-1 more completely, we have used complementary approaches to define the high affinity sites to which it can bind.

SAAB imprinting of oligo 1 with EmBP-1 protein indicates that the three base pairs (positions 4, 3, 2) flanking the ACGT

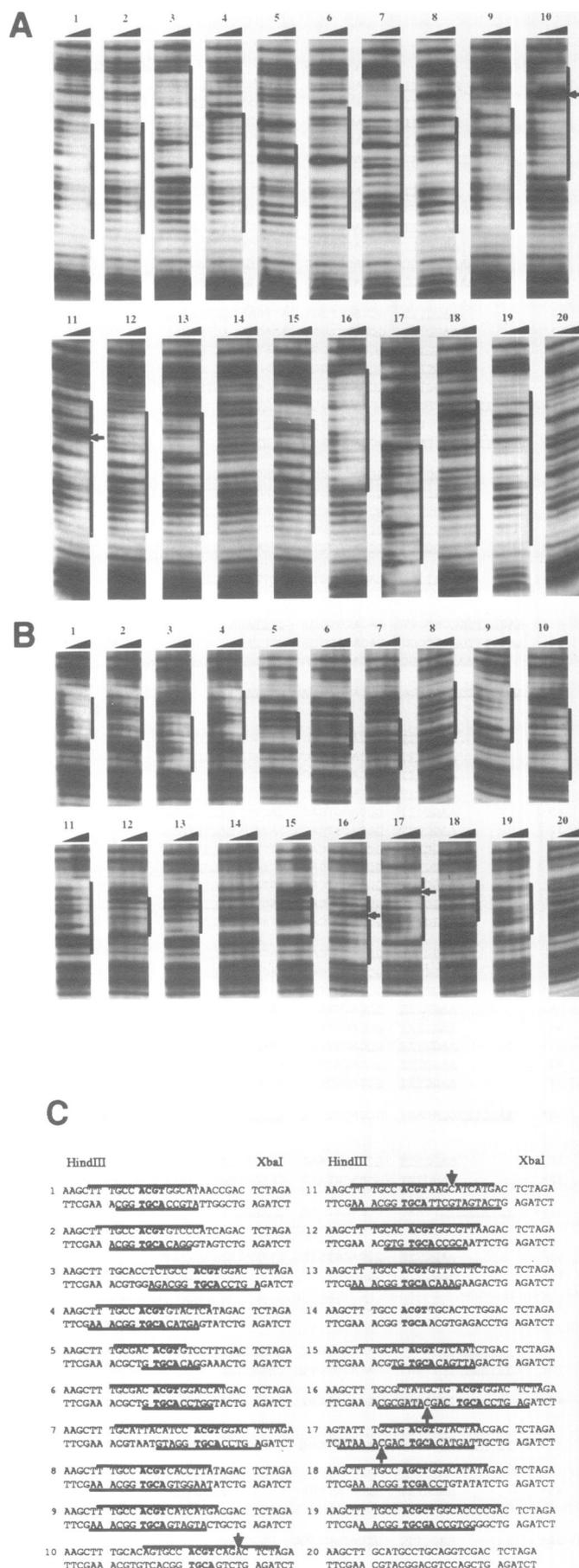
**Table 3.** Classification of the selected binding sites from oligo 2. The base pair changes in other binding sites are not in bold. Symmetric and Dependent half sites are defined according to EmBP-1 binding preference (see text)

		Symmetric	Symmetric	
#		-4-3-2-1-0	+0+1+2+3+4	
1		<b>G C C A C</b>	<b>G T G G C</b>	G-box
2			<b>G T G T C</b>	Em1a
3			<b>G T G G A</b>	G-1A
4			<b>G T G T A</b>	
5		<b>G A C A C</b>	<b>G T G T C</b>	
6			<b>G T G G A</b>	
7		<b>T C C A C</b>	<b>G T G G A</b>	
		Symmetric	Dependent	
#		-4-3-2-1-0	+0+1+2+3+4	
8		<b>G C C A C</b>	<b>G T C A C</b>	Hex
9			<b>G T C A T</b>	
10			<b>G T C A G</b>	
11			<b>G T A A G</b>	
12			<b>G T G T G</b>	
13			<b>G T G T T</b>	
14			<b>G T T G C</b>	
15		<b>G A C A C</b>	<b>G T G T G</b>	
16		<b>T C C A C</b>	<b>G T C A G</b>	
17		<b>T A C A C</b>	<b>G T C A G</b>	
		No ACGT core		
#		-4-3-2-1	+1+2+3+4	
18		<b>G C C A</b>	<b>G C T G G A</b>	
19			<b>C G C T G G C</b>	

**Figure 4.** DNA sequences of sites bound by EmBP-1. Sites were selected from random oligonucleotide pool 2 (Table 1). The selected sites are aligned at ACGT core (except #18 and #19) and grouped according to flanking sequences. Individual clone designations are indicated on the left, r indicates the reverse sequence is presented (relative to flanking sites). The consensus sequences (and number designations) are shown on the right. Sequences representing each of the consensus classes which were used as probes in EMSA and footprinting experiments are indicated with a \*.



**Figure 5.** EMSA analysis of the selected DNA binding sites. (A) Gel shift assay. Equal amounts of His6-EmBP-1a were added to binding reactions containing the DNA probes indicated in Figure 4 (\*). The numbering of DNA binding sites is the same as in Table 3. Sites 1–7 are composed of two symmetric half sites. Sites 8–17 are composed of a symmetric half site and a dependent half site. Sites 18 and 19 do not contain the ACGT core, site 20 is the PUC18 polylinker sequence. Positions of free DNA and protein–DNA complexes are indicated with arrows. (B) Hierarchical affinity of the selected binding sites. A summary of EMSA and DNase I footprinting analysis (see Figure 6) of the selected DNA binding sites in order of affinity with EmBP-1. The numbering of DNA binding sites is the same as in Table 3, Figure 5A. Radioactivity in bound and free fractions from EMSA gels was measured by direct isotope counting. Relative affinity was calculated as the percentage of bound DNA vs. total input DNA, and is normalized to site # 1 in each experiment. The average of three complete experiments are shown as are the standard deviations of the means.



**Figure 6.** DNase I footprinting analysis of selected DNA binding sites. (A) DNase I footprinting on the top strands. The numbering of DNA binding sites is the same as in Table 3, Figure 5A and 5B. Approximately 100 ng, 200 ng and 400 ng of His6-EmBP-1a were analyzed. Protected regions are indicated as a line to the right, hypersensitive sites are indicated with arrows. (B) DNase I footprinting on bottom strand. See (A) for explanation. (C) Summary of DNase I footprinting results. Protected regions are indicated with lines. The nucleotide sequences were determined by Maxam–Gilbert sequencing [41]. The ACGT core and two pseudo ACGT sequences are in bold, hypersensitive sites are indicated with arrows. Oligonucleotide designations are on the left.

core are important in determining interactions between EmBP-1 and its target sequence. Sequences outside position 4 (positions 5 and 6) are also selected weakly, indicating that EmBP-1 interacts with these positions with a moderate degree of sequence preference. From our results, it is evident that the selected binding sites are not symmetric.

All possible binding sites (10) identified from selection of oligo 1 are composed of symmetric half sites and most of them (7) were also cloned from selection of oligo 2. It is consistent that the half site  $G_{-4} C_{-3} C_{-2} A_{-1} C_{-0}$  which is selected strongly in oligo 1 allows for more combinations with other half sites in oligo 2 selection as compared to lower affinity half sites. The selectivity of our system is demonstrated by the fact that the weak symmetric binding site TACACGTGTA is not evident in either of the selections. Contrary to the previous prediction [16], the symmetric binding site GACACGTGTC is not a strong binding site for EmBP-1.

Three lines of evidence support the conclusion that position +4 or -4 affects DNA binding strongly. First, SAAB imprinting of oligo 1 demonstrated strong selection at position +4. Secondly, the sites selected from oligo 2 show similar selection at position -4 and the flanking sequences GC on the left site and GA on the right of the random sequence contributed differentially to the site selection. Finally, evidence from gel shift assays showed that a change of  $G_{+4}$  to  $A_{+4}$  in the highest affinity binding site will lower the binding affinity by approximately 25% (#1 and #4 in Figure 5B). From these data we conclude that at least ten base pairs constitute a strong EmBP-1 binding site, although additional bases flanking the sites might also be important in the interaction.

Most of the sites selected from the random oligo 2 pools contain the ACGT core, consistent with the fact that ACGT is the main component in binding sites of plant bZIP proteins. The highest affinity site for EmBP-1 (GCCACGTGGC) is a perfect palindrome, reflecting the symmetric association of bZIP homodimers with their DNA targets. However, EmBP-1 also binds other non-palindromic sequences with comparable affinity. One high affinity non-symmetric site which was selected is the ABRE, a *cis*-regulatory element involved in abscisic acid mediated gene expression [20]. It is reasonable to assume that such high affinity sites will also be bound *in vivo*, thus the binding specificity of EmBP-1 is consistent with its potential role in mediating ABA regulated gene expression. It is possible that the affinity of EmBP-1 for its various target sites could have biological significance in determining activation efficiency of different genes containing these sites *in vivo* as has been shown for the yeast bZIP protein GCN4 [39].

The optimal binding sites for plant bZIP proteins have two central CG base pairs. EmBP-1 can also bind a target site with three central base pairs with respectable affinity. This situation is also analogous to GCN4 whose optimal site has one central C base pair, and yet it can also bind the central two base pair CG site but with lowered affinity [38]. These facts demonstrate that the bZIP proteins show considerable flexibility in the spatial relationships of the two half sites in recognition sequences and may reflect a degree of flexibility of DNA sequences as shown in the recent crystallographic study of GCN4-CRE complex [40].

EmBP-1 can also bind a large number of asymmetric sites with affinity comparable to the symmetric sites. Some of the asymmetric sites bound by EmBP-1 contain half sites which are not bound if made symmetric (dependent half sites) including the sequences TCA and TAA which are also found in ocs and

as-1 elements used to isolate plant C-box binding bZIP proteins [15,35] and also in binding sites for the yeast bZIP protein GCN4 [36]. This common recognition sequence of different bZIP subfamilies may indicate a common binding mechanism for bZIP-DNA interaction. It is also important in that DNA sequences significantly different from the palindromic G-box could be important regulatory sites *in vivo*. Interestingly, some (but not all) of these sequences show DNase I hypersensitivity upon DNA binding by EmBP-1. This feature could also play a role in the regulatory activity of a protein-DNA complex.

We show that EmBP-1 has a relaxed DNA binding specificity as was noted for the mammalian bZIP protein C/EBP [3]. Thus, it is difficult to generalize 'rules' governing EmBP-1 DNA binding specificity.  $G_{-4} C_{-3} C_{-2} A_{-1}$  is a strong symmetric half site while  $T_{-4} C_{-3} C_{-2} A_{-1}$  and  $T_{-4} A_{-3} C_{-2} A_{-1}$  are weaker symmetric half sites. However, sites composed of any of these symmetric half sites and the dependent half site  $T_{+1} C_{+2} A_{+3} G_{+4}$  are all very high affinity sites (Figure 5B, sites #10, #16 and #17). Although a large number of different sequences were tested, it is likely that these selected binding sites do not represent all of the sequences that EmBP-1 can bind. It is possible that some binding sequences were less abundant due to PCR bias and therefore were not cloned. Accordingly, relative abundance of particular sequences in the selected pool does not necessarily represent relative affinity of the selected sequence.

The wheat bZIP proteins EmBP-1, HBP-1a and HBP-1b can bind Hex-like sequences (GCCACGTAC) with high affinity ([15,31] and this study), however they differ in C-box binding affinity. HBP-1b binds C-box sequence with high affinity, HBP-1a binds the C-box weakly and EmBP-1 has very low C-box binding affinity [15,18]. The basic region of HBP-1b is more similar to that of tobacco TGA1a than it is to HBP1a or EmBP-1. Consistently, TGA1a binds the C-box strongly but it can also bind the Hex element [18]. These observations suggest that DNA binding specificities of plant bZIP proteins range from G-box to C-box and with intermediate DNA binding specificities for sequences such as Hex. Studies of plant bZIP DNA binding specificities in the literature to date have been obtained with purified proteins in their homodimer states. To better understand the role of the bZIP proteins in gene regulation, it will be important to understand how heterodimerization alters the binding preferences of this class of proteins and to develop assay systems to test the combinatorial functions *in vivo*.

## ACKNOWLEDGEMENTS

We would like to thank Drs Jerry Workman and David Gilmore for critical review of this manuscript and Linda Miller for excellent technical assistance. This work was supported in part by grants to MJG from the NSF (MCB 920 6095) and from The Pennsylvania State University, College of Agricultural Sciences (CRIS 3278).

## REFERENCES

1. Pabo, C. O. and Sauer, R. T. (1992) *Annu. Rev. Biochem.*, **61**, 1053-1095.
2. Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988) *Science*, **240**, 1759-1764.
3. Johnson, P. F. and McKnight, S. L. (1989) *Annual Rev. Biochem.*, **58**, 799-839.
4. Neuberger, M., Schuermann, M., Hunter, J. B. and Muller, R. (1989) *Nature*, **338**, 589-590.

5. Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1989) *Science*, **243**, 1681–1688.
6. Gentz, R., Rauscher III, F. J., Abate, C. and Curran, T. (1989) *Science*, **243**, 1695–1699.
7. Agre, P., Johnson, P. F. and McKnight, S. L. (1989) *Science*, **246**, 922–925.
8. Ransone, L. J., Wamsley, P., Morley, K. L. and Verma, I. M. (1990) *Mol. Cell. Biol.*, **10**, 4565–4573.
9. Talanian, R. V., McKnight, C. J. and Kim, P. S. (1990) *Science*, **249**, 769–771.
10. Park, C., Campbell, J. L. and Goddard III, W. A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 9094–9096.
11. Cuenoud, B. and Schepartz, A. (1993a) *Science*, **259**, 510–513.
12. Cuenoud, B. and Schepartz, A. (1993b) *Proc. Natl. Acad. Sci. USA*, **90**, 1154–1159.
13. Johnson, P. F. (1993) *Mol. Cell. Biol.*, **13**, 6919–6930.
14. Foster, R., Izawa, T. and Chua, N.-H. (1994) *FASEB*, **8**, 192–200.
15. Tabata, T., Nakayama, T., Mikami, K. and Iwabuchi, M. (1991) *EMBO J.*, **10**, 1459–1467.
16. Williams, M. E., Foster, R. and Chua, N.-H. (1992) *Plant Cell*, **4**, 485–496.
17. Armstrong, G. A., Weisshaar, B. and Hahlbrock, K. (1992) *Plant Cell*, **4**, 525–537.
18. Izawa, T., Foster, R. and Chua, N.-H. (1993) *J. Mol. Biol.*, **230**, 1131–1144.
19. Guiltinan, M. J., Marcotte, W. R. and Quatrano, R. S. (1990) *Science*, **250**, 267–271.
20. Quatrano, R. S., Guiltinan, M. J. and Marcotte, W. R. (1992) Control of Plant Gene Expression. CRC press, FL, 69–90.
21. Oliphant, A. R., Brandl, C. J. and Struhl, K. (1989) *Mol. Cell. Biol.*, **9**, 2944–2949.
22. Gubler, U. and Hoffman, B. J. (1983) *Gene*, **25**, 263–269.
23. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor.
24. Van Dyke, M. V., Sirito, M. and Sawadogo, M. (1992) *Gene*, **111**, 99–104.
25. Guiltinan, M. J. and Miller, L. (1994) *Plant Mol. Biol.*, in press.
26. Blackwell, T. K. and Weintraub, H. (1990) *Science*, **250**, 1104–1110.
27. Devos, K. M., Atkinson, M. D., Chinoy, C. N., Guiltinan, M. J., Quatrano, R. S. and Gale, M. D. (1991) *Theoret. Appl. Genet.*, **82**, 665–667.
28. Blochlinger, K., Bodmer, R., Jack, J., Jan, J. Y. and Jan, Y. N. (1988) *Nature*, **333**, 629–635.
29. Biggin, M. D., Bickel, S., Benson, M., Pirotta, V. and Tjian, R. (1988) *Cell*, **53**, 713–722.
30. Mitchell, P. J. and Tjian, R. (1989) *Science*, **245**.
31. Tabata, T., Takase, H., Takayama, S., Mikami, K., Nakatsuka, A., Kawata, T., Nakayama, T. and Iwabuchi, M. (1989) *Science*, **245**, 965–967.
32. Schindler, U., Menkens, A. E., Beckmann, H., Ecker, J. R. and Cashmore, A. R. (1992) *EMBO J.*, **11**, 1261–1273.
33. Meier, I. and Grisse, W. (1994) *Nucl. Acids Res.*, **22**, 470–478.
34. Katagiri, F., Lam, E. and Chua, N.-H. (1989) *Nature*, **340**, 727–730.
35. Singh, K., Dennis, E. S., Ellis, J. G., Llewellyn, D. J., Tokuhisa, J. G., Wahleither, J. A. and Peacock, W.J. (1990) *Plant Cell*, **2**, 891–903.
36. Mavrothalassitis, G., Beal, G. and Papas, T. S. (1990) *DNA and Cell Biol.*, **9**, 783–788.
37. Fromm, H., Katagiri, F. and Chua, N.-H. (1991) *Mol. Gen. Genet.*, **229**, 181–188.
38. Sellers, J. W., Vincent, A. C. and Struhl, K. (1990) *Mol. Cell. Biol.*, **10**, 5077–5086.
39. Hill, D. E., Hope, I. A., Macke, J. P. and Struhl, K. (1986) *Science*, **234**, 451–457.
40. König, P. and Richmond, T. J. (1993) *J. Mol. Biol.*, **223**, 139–154.
41. Maxam, A. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499–560.