Plant Molecular Biology **26**: 1041–1053, 1994.

Molecular characterization of the DNA-binding and dimerization domains of the bZIP transcription factor, EmBP-1

Mark J. Guiltinan* and Linda Miller
*The Pennsylvania State University, Department of Horticulture, The Biotechnology Institute, Center for Gene Expression, University Park, PA 16802, USA (*author for correspondence)*

Received 30 April 1994; accepted in revised form 15 August 1994

Key words: Abscisic acid, bZIP proteins, DNA-binding, EmBP-1, leucine zipper, transcription factor

Abstract

The wheat basic-leucine zipper (bZIP) DNA-binding protein EmBP-1 has been implicated in the mechanisms of abscisic acid (ABA) mediated gene regulation. Sequence and structural homology to the yeast bZIP protein GCN4 has been used to predict the location of the functional domains of EmBP-1. In order to test these predictions, the presumptive DNA-binding and dimerization domains of EmBP-1 were mapped by producing a series of truncated protein fragments and functionally testing them *in vitro*. Deletion of 5 amino acids of the predicted basic domain resulted in a loss of all DNA-binding activity. A fragment containing all six leucine repeat elements showed strong DNA-binding activity. Sequential deletion of the leucine repeat elements resulted in first an increase in DNA-binding activity (–L6 and –L5) followed by a reduction in binding activity (–L4) and eventually complete elimination of all detectable DNA-binding activity (–L3 and –L2). This demonstrates the importance of an intact leucine zipper domain of at least 4 repeat elements for efficient DNA-binding. The smallest polypeptide that retained DNA-binding activity is a fragment spanning amino acid residues 248–308 (ca. 8.4 kDa) consisting of minimal basic and leucine zipper domains. Dimerization of EmBP-1 was demonstrated by co-translation of fragments of differing molecular weights and identification of a DNA-protein complex with intermediate mobility to that produced by each fragment alone. A unique leucine-proline repeat element found N-terminal to the DNA-binding domain of EmBP-1 does not appear to play a role in DNA-binding or dimerization. These results confirm the locations of the functional domains of EmBP-1 predicted by similarity to GCN4. The high degree of functional conservation of the bZIP proteins spanning organisms from plants to fungi highlights the ancient origin of this class of transcription factors and of the mechanisms of gene regulation in which they participate.

Introduction

ABA is a plant hormone involved in stomatal closure, the induction and maintenance of seed maturation, root shoot signaling and in various plant responses including osmotic and pathogenesis-related stress [29]. During late seed development in many plant species, a set of genes is induced in response to elevated levels of ABA [29]. These include the wheat Em gene (embryo maturation), which encodes a small hydrophilic protein thought to be involved in protection of the

embryo during the extreme dehydration of dormancy. The promoter region of the wheat Em gene and other ABA-induced genes contain *cis*-acting DNA sequences (ABA-responsive element or ABRE) which are essential to ABA-mediated transcriptional enhancement as defined by both gain and loss of function gene-fusion experiments [10, 19–21, 26, 36, 39]. Mutagenesis of two base pairs of the Em ABRE eliminates its ability to confer ABA responsiveness on a viral promoter in a transient gene expression assay [10].

DNA-binding proteins have been identified in crude nuclear extracts from rice and wheat cells which interact with the Em ABRE in *in vitro* binding assays [10]. cDNAs encoding Em ABRE-binding proteins (EmBPs) were isolated by screening a wheat embryo λgt11 expression library for phage expressing proteins capable of binding a synthetic oligonucleotide probe containing the ABRE sequence [10]. These cDNAs encode proteins with binding properties indistinguishable from those observed in the nuclear extracts as judged by competitive DNA-binding analysis and methylation interference footprinting. Molecular characterization of the EmBP cDNAs revealed that they encode a family of closely related DNA-binding proteins (the EmBP-1 family) of the (bZIP) class of transcription factors. Molecular-genetic mapping indicated that there are at least 7 closely related genes in the wheat EmBP-1 family [4]. The same 2 bp mutant ABRE sequence (mentioned above) which is not ABA- responsive in transient assays is not recognized by EmBP-1 [10]. The DNA-binding specificity of EmBP-1 and its relatedness to a well known class of transcriptional activating proteins implicates it in the mechanisms of ABA-mediated gene activation.

Members of the bZIP class of transcription factors have been identified in organisms ranging from yeast to man and have been shown to be involved in the induction of gene expression in a number of systems [5, 11, 15, 33]. The bZIP proteins share two highly conserved regions, the basic and the leucine heptad repeat domains. The structures and functions of the bZIP domains of the yeast regulatory protein GCN4 have been characterized by deletion analysis, mutagenesis, domain swapping and X-ray crystallographic studies [1, 12, 13, 7, 23]. The basic domain, which consists of ca. 26 amino acid residues, is found directly N-terminal to the leucine heptad repeat domain. Between 3 and 6 leucine repeats are found in various bZIP proteins. Two of the amino acid residues of the basic domain (residues N-234 and R-242, Fig. 2A) and each of the leucines of the repeat elements (which are sometimes substituted with methionine, valine, isoleucine or alanine) are strictly conserved between the various bZIP proteins. However, comparisons of the remaining residues between different proteins in these domains generally show conservative substitutions, with mostly basic residues in the basic domain and a discrete pattern of hydrophilic and hydrophobic residues interspersed between the leucine repeats. The crystal structure of GCN4 indicates that the two domains form a slightly curved, continuous coiled coil structure when complexed with DNA [7, 23]. The basic domain contacts the DNA in the major groove while the leucine repeat element forms an amphipathic alpha helix, forming a very efficient dimerization interface. Thus, the bZIP proteins bind DNA as a dimer, with one basic domain from each subunit tracking in the major groove of the recognition DNA sequence.

To date, at least 19 bZIP proteins have been characterized by cDNA cloning from plant species [6, 8, 10, 16, 17, 25, 28, 31, 34, 35, 37, 38, 43, 46]. Interestingly, each of these proteins recognizes DNA sequences that share the central core sequence ACGT. The bases flanking the core sequences are important to the binding affinity of these proteins, presumably through interactions with the basic domain, by influencing initial binding or by altering the structure of the DNA itself. EmBP-1 is a member of one subclass of the plant bZIP proteins that have been called the G-box-binding proteins (or GBFs) for their strong binding preference to the sequence core CACGTG [5, 14, 25, 31–33, 43, 44]. These plant bZIP proteins have been shown to bind to regulatory elements from a wide variety of inducible plant genes including those regulated by cell cycle, light, UV

light, drought and pathogen infections [5, 10, 16, 25, 37, 43]. It is likely that hetero- and homodimerization, cell specific expression and DNA-binding specificity of the bZIP family of proteins combine to determine the functional repertoire of bZIP dimers found in a given cell, and that their interaction with other transcription factors results in regulated gene expression. In order to elucidate the role of the bZIP proteins in regulating plant gene expression, it is first necessary to characterize the dimerization and DNA-binding specificity of each family member.

We have undertaken a detailed characterization of the structure-function relationships of EmBP-1. Sequence and structural homology to the yeast bZIP protein GCN4 has been used to predict the location of the functional domains of EmBP-1. In order to test these predictions, the presumptive DNA-binding and dimerization domains of EmBP-1 were mapped by testing a series of progressive N- and C-terminal deletions for DNA-binding and dimerization activities. In addition, EmBP-1 contains a unique proline interspersed leucine repeat element not found in any other bZIP protein described to date [10]. We have investigated the influence of this leu-pro domain on the DNA-binding activity of EmBP-1. We show that the structure and location of the DNA-binding and dimerization domains of EmBP-1 correspond closely to those found in GCN4, demonstrating the conservation of these domains between higher plants and yeast and the ancient origin of this class of transcription factors.

Materials and methods

PCR-based coupled transcription/translation (TNT)

DNA coding for the truncated versions of EmBP-1 was generated using polymerase chain reactions (PCR) with the GC19 *Eco* RI fragment as a template [10]. This fragment spans amino acids 150 to 338 of the full-length protein coded for on the cDNA pGCNN546 (GenBank accession number UO7933). The 5′ primer employed in the polymerase chain reaction carried a T7 promoter and an in-frame ATG codon. Figure 1 outlines the protocol for the PCR-based coupled transcription/translation reactions and describes the basic design of the 5′ and the 3′ primers. Table 1 lists the nucleotide sequences of these primers. The polymerase chain reaction was in a volume of 100 μl containing the following components: 3 ng of template DNA, 20 pmol of primers, and 100 μmol of deoxynucleotide triphosphates. The cycling parameters were denaturation at 95 °C for 1.0 min, annealing for 1.0 min, and extension at 72 °C for 1.5 min. The cycle was repeated 35 times. The annealing temperature was optimized for each set of primers. Proteins were produced for electrophoretic mobility shift assay (EMSA) by transcribing 1 μg of template with T7 RNA polymerase followed by *in vitro* translation in 50 μl of reticulocyte lysate in the presence of ^{35}S-methionine (TNT reaction) according to the recommendations of the manufacturer (Promega).

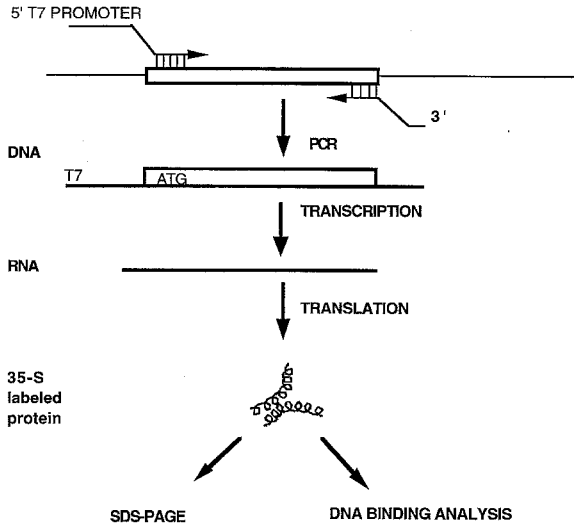

Fig. 1. Schematic depiction of the PCR-based coupled transcription/translation system used to generate full-length and deletion fragments of EmBP-1 for use in *in vitro* DNA-binding analysis. Each 5′ primer consists of an *Eco* RI restriction enzyme site, a T7 RNA polymerase promoter, an ATG, and the appropriate nucleotide sequence coding for the initial seven amino acids of the 5′ end of the protein. The 3′ primer consists of an *Eco* RI restriction enzyme site, one stop codon, and the appropriate nucleotide sequence coding for the terminal seven amino acids of the respective fragment.

Table 1. Nucleotide sequences of primers used in the PCR based TNT system. *Eco* RI restriction sites are shown in lower case. Underlined nucleotides code for the T7 RNA polymerase promoter. ATG start codons and stop codons are in bold letters. Immediately following the start or stop codons are the EmBP-1 homologous sequences used for hybrid formation, except for in C-terminal primer 338 which is homologous to the sequence flanking the *Eco* RI site of λgt11 and was used to amplify the original λGC19 partial cDNA encoding EmBP-1 [10].

Terminal amino acid	Sequence 5′ to 3′
N-terminal	N-terminal T7 primers
150	ggaattcTAATACGACTCACTATAGGGAGCC**ATG**GCCAATCAGAAGGGCTCATCA
200	ggaattcTAATACGACTCACTATAGGGAGCC**ATG**GTTTTGGCACCTGGGAGGGCG
248	ggaattcTAATACGACTCACTATAGGGAGCC**ATG**GATGAACGGGAACTGAAGAGG
260	ggaattcTAATACGACTCACTATAGGGAGCC**ATG**AAACAATCCAACAGAGAGTC
C-terminal	C-terminal primers
317	gaattcTCA**TTA**CATCAGCTGTTTATTTTCTGT
338	GGG**TCACTCAGTCA**TTGACACCAGACCAACTGGTAATG
315	ggaattc**TCA**CTGTTTATTTTCTGTTTCCAT
308	ggaattc**TCA**GGTTTTGCAGTCCTTCTTAAG
301	ggaattc**TCA**CTGGTCGAGTTCTGATCTGAG
294	ggaattc**TCA**CGTGCCGTTCGCTGCGGTCAG
287	ggaattc**TCA**CTCACTTACCTTCTGGGCTAG

Characterization of ^{35}S-labelled protein

To verify integrity and to quantitate the amount of protein produced *in vitro*, a portion of the TNT reaction was applied to a denaturing 10% tricine glycine gel (denaturing PAG) and electrophoresed [30]. The gels were dried, autoradiographed and direct measurement of the ^{35}S in the protein bands was performed by Betascope analysis (Betascope Model 603, Betagen Corp.). Based on this quantitation, the volumes of the TNT reactions that would contain equimolar amounts of labelled proteins were calculated taking into account the number of methionines in each fragment.

Binding reactions and EMSA

Binding reactions were performed according to Guiltinan *et al.* [10] in a final volume of 10 μl containing the following components: 24 mM Tris pH 7.9, 24% glycerol, 70 mM KCl, 0.14 mM EDTA, 2.15 mM DTT, 15 mM MgCl2, 500 ng poly-dIdC, and 100 pg of 32P-end-labelled abscisic acid-response element (ABRE). The DNA fragment containing the ABRE was derived by a *Hind* III digestion of pMG 76.11 [10]. The binding reaction was initiated by adding the appropriate amount of ^{35}S-labelled protein generated in the TNT reactions. When adding equimolar amounts of protein fragments from a deletion series to a set of binding reactions, the amount of TNT reaction was adjusted (with unprogrammed TNT reaction mixture) so that a constant volume of reticulocyte lysate was added. After incubation for 10 min at room temperature, the binding reaction was applied to a 4% acrylamide (40:1 acrylamide bisacrylamide) gel in high-ionic-strength electrophoresis buffer (25 mM Tris pH 8.9, 190 mM glycine, 1 mM EDTA). Before applying the binding reaction, the gel was pre-electrophoresed in the above buffer for one hour at 150 V at 4 °C. After the sample was applied, the gel was electrophoresed for 2 h under the same conditions. Gels were dried and autoradiographed on Kodak XAR 5 film at −80 °C with 2 Lightning Plus (Dupont) intensifying screens.

Molecular weight determination of proteins in DNA complexes

The sizes of proteins in DNA/protein complexes were determined by two-dimensional gel electro-

phoresis performed according to the method of Hope and Struhl [12]. An EMSA with the appropriate protein fragment and ABRE was performed as described above except that the gels were not dried. After electrophoresis, the gel was autoradiographed at 4 °C and the appropriate acrylamide fragments, containing DNA and protein, were excised. These excised fragments were inserted into the well of a denaturing 10% tricine glycine gel and electrophoresed. The gels were fixed, treated with Entensify (NEN), dried and autoradiographed.

Results and discussion

In vitro *expression system*

In order to map the dimerization and DNA-binding domains of EmBP-1, we have used an *in vitro* expression system (Fig. 1) which allows for the rapid production and assay of various subfragments of EmBP-1 protein. 5′ PCR primers were designed (see Table 1) which include a cloning site, signals for transcription (T7 polymerase promoter), an ATG codon in frame with the designated codon of EmBP-1 and a 21 bp region of homology to EmBP-1. The 3′ primers consists of 21 bp of EmBP-1 homologous sequence followed by a termination codon and cloning site. After amplification via PCR, the DNA products were used in coupled transcription-translation reactions as described above. In all cases, polypeptides of the expected molecular weights were detected, but, in some cases minor amounts of products of smaller molecular weight were also observed (for example, see Fig. 3B).

Amino-terminal deletions

In order to test the role of the leucine proline repeat element N-terminal to the basic domain and to determine the N-terminal boundary of the DNA-binding domain of EmBP-1, a series of protein fragments were synthesized with progressive N-terminal deletions (Fig. 2A). In the case of the +L6, and +L/P fragments, multiple protein/DNA complexes were seen with EMSA as opposed to the −L/P fragment which produced a single band (Fig. 2B). These complexes were shown to represent specific protein interactions by competition experiments with unlabelled specific and non-specific competitor DNAs (data not shown) and to be dependent on the inclusion of EmBP-1 protein by the lack of DNA-binding in an unprogrammed reticulocyte transcription/translation control (Fig 2B, −). A further deletion of 12 amino acids (to amino acid 260) resulted in the complete elimination of all DNA-binding activity (Fig. 2B, −BD fragment).

By comparison with the amino acid sequence of the yeast bZIP protein GCN4, the EmBP-1 −BD fragment has a deletion of the first five amino acids of the predicted basic domain (Fig. 2A). The elimination of DNA-binding by the removal of these residues is consistent with the predicted location of the EmBP-1 basic domain and its presumptive function in DNA-binding. It is interesting to note, that even though the crystal structure of GCN4 predicts that the amino acids removed in −BD do not directly contact DNA, these results indicate that they are essential for DNA-binding [7, 23]. This indicates that the DNA-binding domain of EmBP-1 is dependent on this portion of the basic domain for proper recognition and binding to its target sequences, perhaps by a regional effect on the structural conformation of the basic domain.

The leucine proline repeat element N-terminal to the basic domain is deleted in the protein fragment −L/P yet this fragment retains DNA-binding activity (Fig. 2B, −LP). This indicates that the leucine-proline domain unique to EmBP-1 is not necessary for DNA-binding. However, the −L/P protein/DNA complex runs as a single band on an EMSA in contrast to the +L/P fragment which produced three complexes (Fig. 2B, +LP). One explanation for the production of multiple DNA/protein complexes of differing mobilities is that there is a conformational difference in the protein fragment containing the leucine proline repeat element that is not present in the −L/P fragment. A second possible explanation would

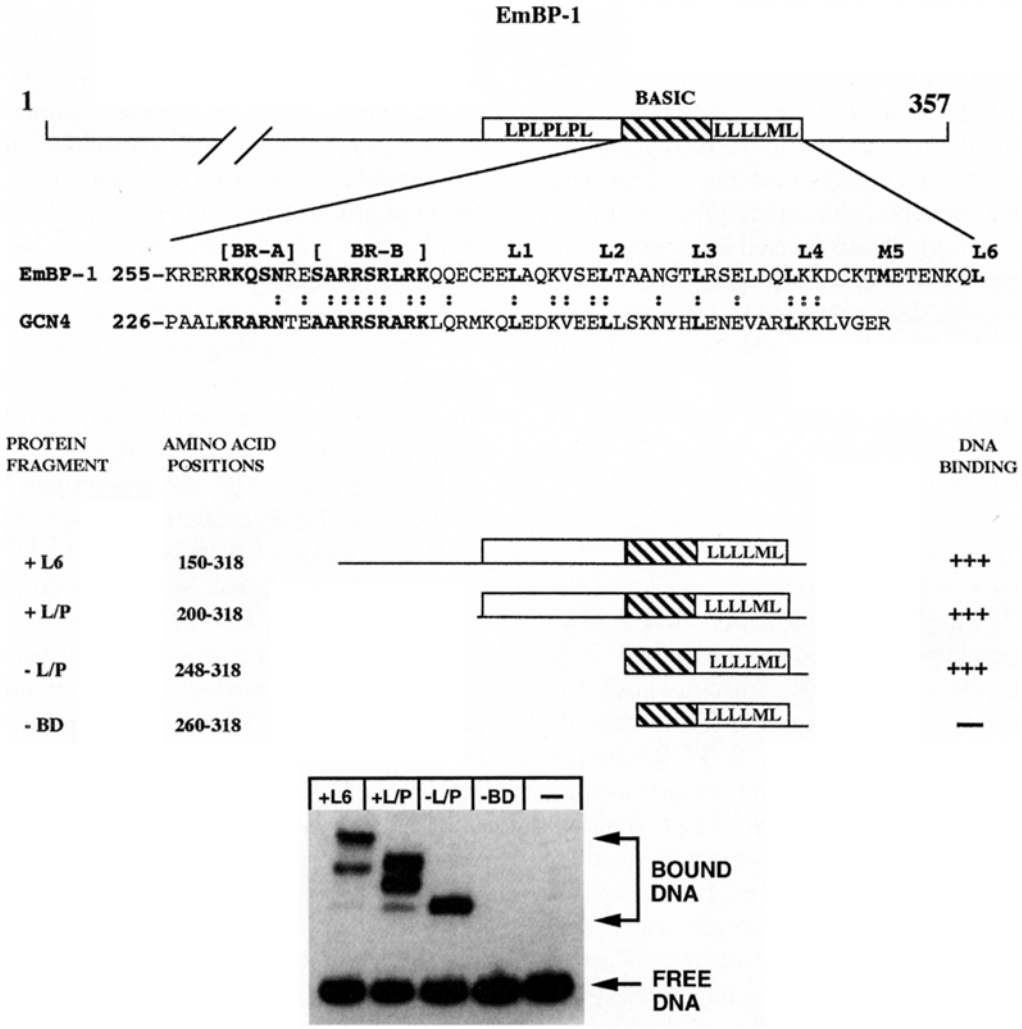

Fig. 2. DNA-binding analysis of amino-terminal deletions of EmBP-1. A (top). Map of EmBP-1 and truncated versions of the protein generated in PCR-based TNT reactions. Amino acid sequence alignment of the EmBP-1 and GCN4 bZIP domains. Conserved amino acids are indicated in bold letters, and conserved basic region subdomains BR-A and BR-B are indicated according to Vinson *et al.* [40]. Positions of the conserved basic and leucine zipper domains and of the leucine proline repeats are shown. The end-points of the full-length protein and the truncated fragments and their designations are indicated to the left. Relative DNA-binding activities are indicated on the right. B (bottom). Autoradiograms of EMSA for each protein fragment as well as unprogrammed reticulocyte lysate (–). Equimolar quantities of the ^{35}S-labelled proteins were added to each binding reaction. Each deletion is indicated by the labels designated in A. Arrows indicate the positions of the unbound DNA and the protein-DNA complex.

be the formation of multimeric protein/DNA complexes. Alternatively, a mobility difference could result from protein/DNA complexes containing proteins of differing molecular weights. As mentioned previously, minor amounts of partial products were occasionally evident in the autoradiograms of the TNT reactions, and these could be the cause of the multiple bands. Therefore, to distinguish between these possibilities, the molecular weights of the proteins in each of the complexed forms were determined.

Separate acrylamide fragments containing each of the DNA/protein complexes formed with the +LP fragment were excised from an EMSA gel

(Fig. 3A) and were applied to a denaturing tricine glycine gel (Fig. 3B). In order to compare the sizes of the proteins in each of the complexes with the proteins produced in the TNT reaction, an unfractionated TNT reaction was run in an adjacent lane (Fig. 3B, TNT). Two specific detectable polypeptides were synthesized in this reaction, a major protein of the expected size (16 kDa) and a minor protein of lower molecular weight (7 kDa). Additionally, an endogenous hemoglobin (Hb) product was observed in all TNT reactions including unprogrammed controls. The minor 7kDa polypeptide could be the result of either internal translational initiation on the EmBP-1 transcript, premature transcriptional or translational termination or proteolytic cleavage of the full-length protein product. The slower mobility protein/DNA complex seen in the native gel (complex 1, Fig. 3A) contained the 16 kDa 'full-length' protein fragment (Fig. 3B, lane 1) which migrated at the same mobility as the major protein produced in the TNT reaction. Complex 2 contained a 16 kDa protein and the smaller 7 kDa partial product (Fig. 3B, lane 2). This indicates the formation of heterodimers between these two fragments. Complex 3 contains only the 7 kDa protein, most likely due to homo-dimerization of the 7 kDa partial fragment.

These results indicate that the multiple retarded bands of protein/DNA complexes 1, 2 and 3 seen with the longer versions of EmBP-1 are due to the binding of full-length homodimers, heterodimers of full-length and partial products and partial product homodimers respectively. Therefore, the multiple bands produced in an EMSA with the protein fragment including the leucine-proline repeat region are not due to a conformational difference imposed by this element but are due to full-length and partial protein products causing homo- and heterodimers. Even though minor amounts of the partial-length proteins are produced in the TNT reaction, they contribute a significant amount of DNA-binding activity in the reaction. It is possible that the shorter fragments dimerize more efficiently than do the longer EmBP-1 fragments and thus they contribute disproportionately to DNA-binding.

As mentioned above, it is possible that a protease present in the reticulocyte extracts cleaves the 16 kDa EmBP-1 product producing the partial-length 7 kDa product. Indeed, similar proteolytic cleavage has been reported in the specific cleavage of basic helix-loop-helix class transcription factors by a thiol protease, m-calpain [42]. This cleavage results in smaller fragments capable of DNA-binding and dimerization but not tran-

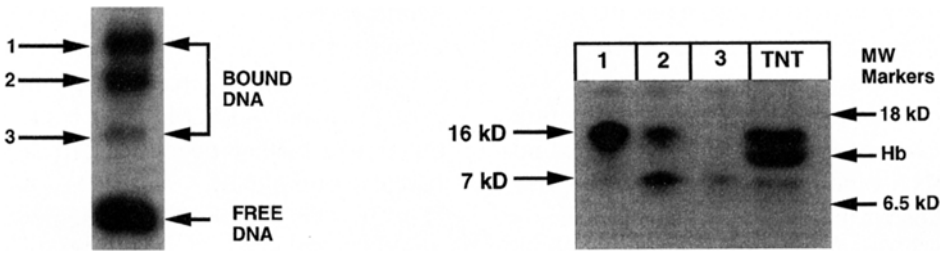

Fig. 3. Two-dimensional gel electrophoresis of protein-DNA complexes. A (left). Autoradiogram of an EMSA with a binding reaction containing a ^{35}S-protein fragment of EmBP-1 containing amino acids #200–318. Each of the protein/DNA complexes are labelled sequentially. An adjacent lane (not shown) was run with an identical binding reaction except for the DNA probe which was not labelled with ^{32}P. The bands visualized in the binding reaction with the labelled probe were used as guides to locate the bands in the EMSA with the unlabelled probe. The complexes in the unlabelled lane were excised for further analysis on a 10% tricine glycine polyacrylamide gel. B (right). Autoradiogram of ^{35}S-labelled proteins in each of the excised protein/DNA complexes separated on a 10% tricine glycine polyacrylamide gel. A TNT reaction containing the unreacted 35-S labelled protein fragment is shown (TNT). Arrows indicate the positions of the EmBP-1 protein fragments and of hemoglobin labelled during the TNT reaction (Hb). Lanes 1–3 contain the protein/DNA complexes that are represented by the bands labelled 1–3 in A. The source of the excised bands was the lane containing a binding reaction with unlabelled probe run adjacent to that shown in A. Positions of molecular weight markers are indicated on the right.

scriptional activation. The authors suggest that this mechanism may be involved in the control of protein turnover and/or functional activity *in vivo*. Interestingly, native nuclear extracts from wheat, rice, cauliflower, tobacco, and *Arabidopsis* all form multiple DNA protein complexes with sequences related to the wheat ABRE upon EMSA [10, 25, 32, 44]. Although neither the nature of these forms nor their biological significance have yet to be elucidated, it is possible that they could result from hetero-dimerization of different bZIP proteins and/or interactions with proteolytic cleavage fragments.

Carboxy-terminal deletions

To test the contribution of the EmBP-1 leucine zipper repeats to DNA-binding, a series of carboxy deletions were designed to remove the leucine repeats sequentially (Fig. 4A). The deletions were constructed so that each fragment included the amino acid sequence up to but excluding each of the leucine residues. A partial length protein spanning the residues encoded in the original λgt11 isolate GC19 [10] and C-terminal deletions were tested for DNA-binding activity by EMSA (Fig. 4B). Relative DNA-binding activities from replicate experiments were quantified by direct radioactivity counting and are presented relative to GC19-binding activity levels (Fig. 4C). The GC19 protein and a fragment containing all six leucine repeat elements showed strong DNA-binding activity. Sequential deletion of the first and second leucine repeat elements produced an increase in DNA-binding (–L6 and –L5) relative to GC-19. Subsequent deletion of 7 amino acids resulted in greatly reduced DNA-binding activity (–L4). Further deletion of repeats 3 and 2 completely eliminated detectable DNA-binding activity (–L3 and –L2). These results clearly show the importance of an intact leucine zipper domain of at least 3 repeat elements for DNA-binding *in vitro*.

These results also raise an intriguing question regarding the functional significance of the two C-terminal repeats expendable for DNA-binding but which have been conserved during evolution of the EmBP-1 leucine zipper. The specificity of DNA-binding has been determined to reside in the basic domain [1]. However, there have been recent reports of a dimerization domain that exhibits separate regions responsible for either homo-dimerization or hetero-dimerization [22]. This hetero-dimerization domain allows for interaction of accessory proteins and suggests a method of conferring further specificity to DNA-binding. In fact, hetero-dimerization has been shown to increase DNA-binding affinities for a specific element or confer specificity for a specific element not seen with the homodimer [3, 9, 15, 18, 24, 45, 47]. The possibility exists that although only the first four leucine repeats in EmBP-1 are necessary for homo-dimerization, the other two repeats may be involved in hetero-dimerization with accessory proteins.

Minimal domain

The results from the N- and C-terminal deletion experiments have independently defined the domains required for strong *in vitro* DNA-binding activity. In order to define the minimal protein fragment capable of binding DNA, a series of C-terminal deletions was tested using the smallest N-terminal deletion fragment designated Minimal N-Terminal (MNT, Fig. 5a). EMSA analysis showed that several of these deletion fragments retained strong DNA-binding activity (Fig. 5b, MNT +L6, MNT –L6 and MNT –M5) but that a further deletion of the fourth leucine repeat completely abolishes DNA-binding (Fig. 5b, MNT –L4). Thus, fragment MNT –M5 (amino acid residues 248–308) is the smallest polypeptide tested that binds DNA. This protein is approximately 8.4 kDa and consists of the basic region and leucine zipper domain up to but excluding the leucine repeat at position Met-5. These results agree with those seen with GCN4 which requires four leucine repeat elements for full DNA-binding activity [11, 13]. However, this contrasts with the results depicted in Fig. 4b where deletion of leucine repeat 4 reduced but did

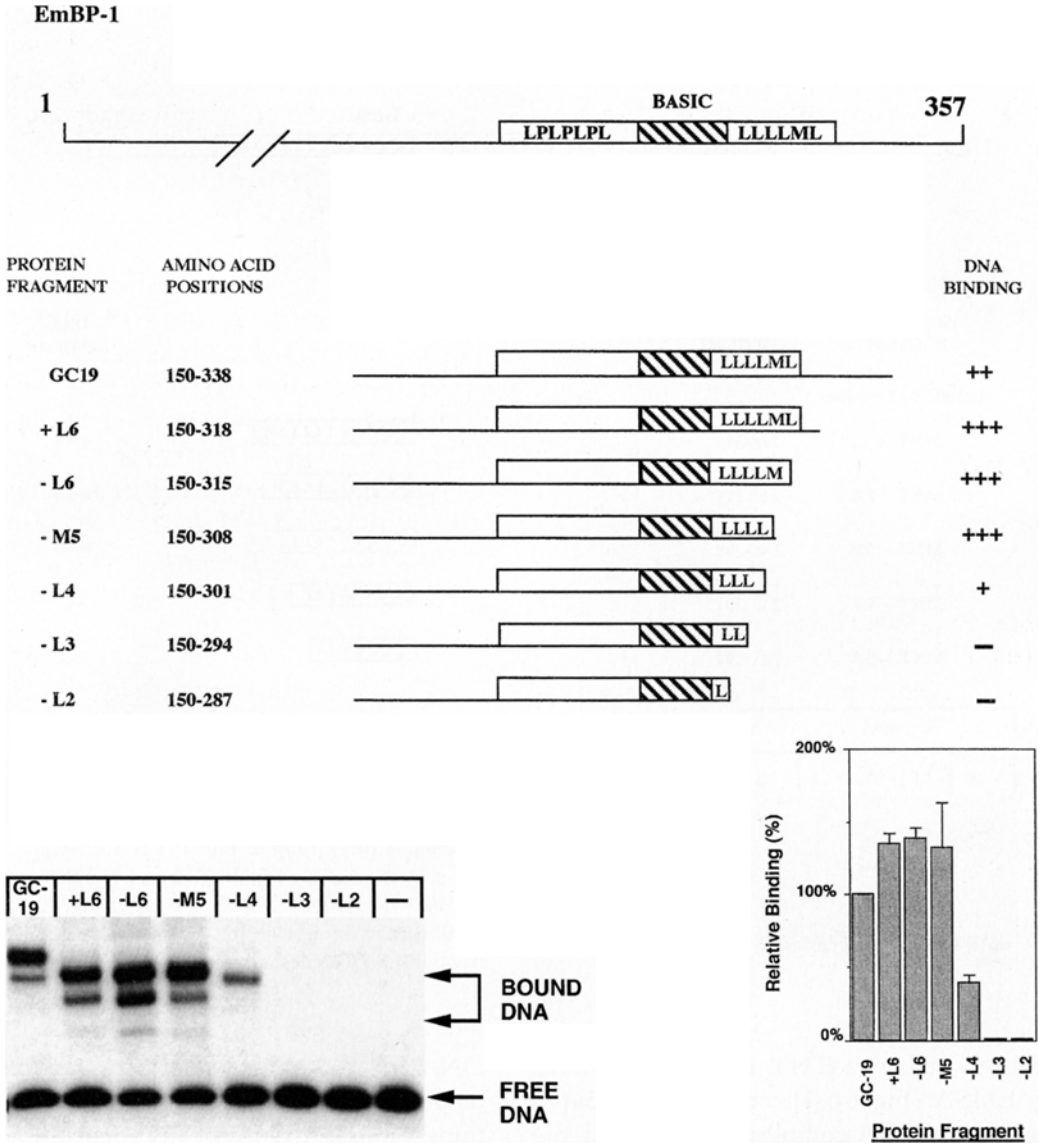

Fig. 4. DNA-binding analysis of carboxy-terminal deletions of EmBP-1. A (top). Map of EmBP-1 and truncated versions of the protein generated in PCR based TNT reactions. Labelling is as in Fig. 2a. B (bottom left). DNA-binding analysis of each protein fragment as well as unprogrammed reticulocyte lysate (–). Binding reactions were as described in Fig. 2b. C (bottom right). Quantitation of DNA-binding activity for protein fragments GC-19 to –L3 of EmBP-1. EMSA analysis was performed with equimolar quantities of protein as shown in Fig. 2B. The radioactivity in the protein/DNA complexes was quantified by Betascope analysis. The results are presented relative to the binding activity of the GC-19 protein (fold binding). Results are displayed as the mean with error bars showing standard deviations from at least three independent experiments for each protein fragment.

not abolish DNA-binding activity. This could be due to further stabilization of the dimerization or DNA-binding regions by residues N-terminal of the basic domain.

Dimerization of EmBP-1

To demonstrate dimerization of EmBP-1, two protein fragments of differing molecular weights

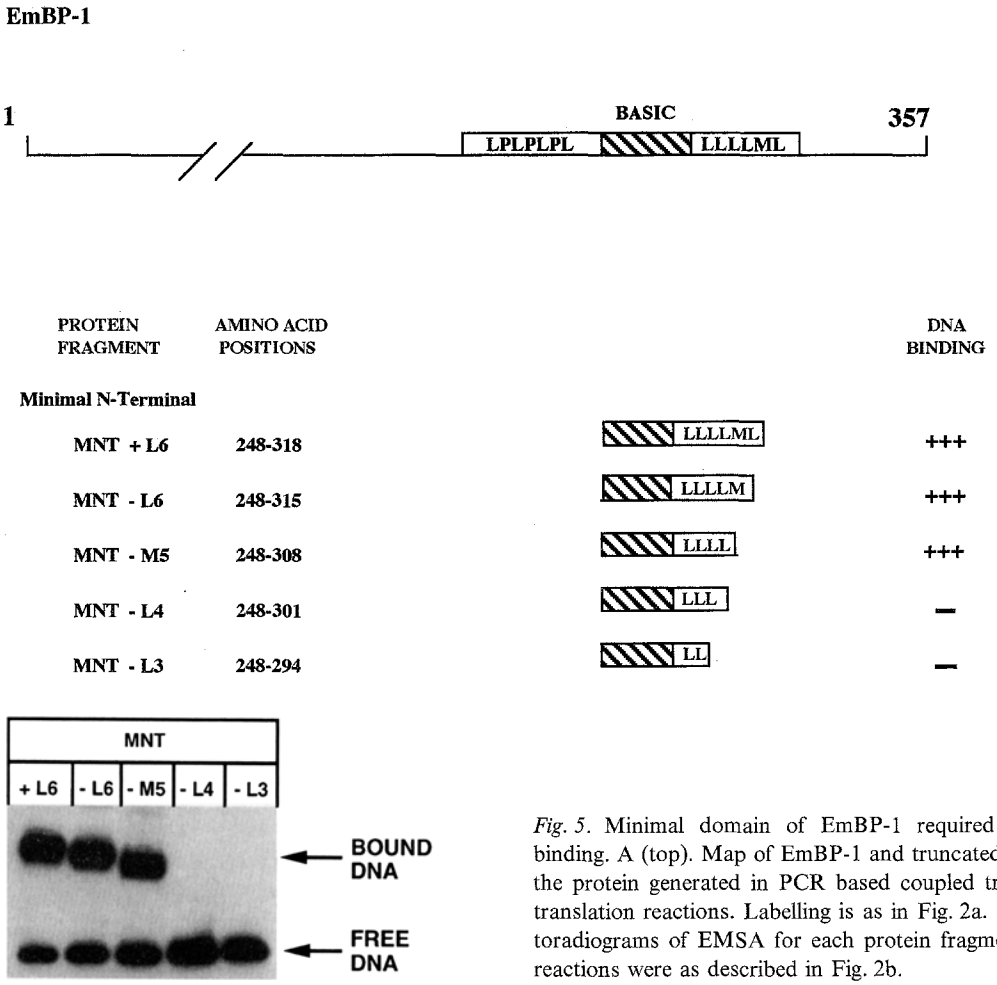

Fig. 5. Minimal domain of EmBP-1 required for DNA-binding. A (top). Map of EmBP-1 and truncated versions of the protein generated in PCR based coupled transcription/translation reactions. Labelling is as in Fig. 2a. B (left). Autoradiograms of EMSA for each protein fragment. Binding reactions were as described in Fig. 2b.

were co-translated in a TNT reaction and assayed by EMSA (Fig. 6). Theoretically, the mobility of a protein/DNA complex is determined in part by and is inversely proportional to the molecular weight of the protein fragment. Consequently, if two protein fragments of differing molecular weights are co-translated, a heterodimer consisting of the two proteins would form a protein/DNA complex migrating at a mobility intermediate to the homodimer protein/DNA complexes. To simplify interpretation of the results, protein fragments that produced a single band in an EMSA were chosen (amino acid residues 248–338 and 248–308).

The difference in size between these two proteins was enough to produce homodimer protein/DNA complexes that migrated at different mobilities (Fig. 6 lanes 1 and 6, individual translations). A series of TNT reactions were devised so that the amount of input DNA coding for the larger protein fragment remained constant while the DNA coding for the smaller protein fragment gradually increased (Fig. 6, lanes 2–5). Binding reactions with the co-translated protein fragments resulted in three protein/DNA complexes representing each of the homodimers alone and a complex that migrated at an intermediate mobility and only appeared in the binding reactions using co-translated proteins. These results show that EmBP-1 binds DNA as a dimer, a prediction previously made based on amino acid similarity with other bZIP proteins [2, 12, 27, 41].

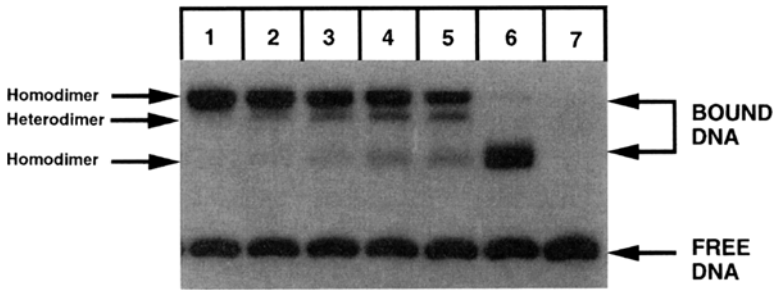

Fig. 6. EmBP-1 binds DNA as a dimer. EMSA of DNA-binding reactions containing labelled ABRE probe with protein fragments translated either individually from equal amounts of DNA template (lanes 1 and 6) or cotranslated (lanes 2–5). Reactions run in lanes 1–5 contained identical loadings of a fragment spanning residues 248–338. Reactions run in lanes 2–6 contained increasing amounts of a protein fragment spanning residues 248–308 produced by diluting the DNA template added to the TNT reaction in a two fold series (lane 6, 1–, lane 5, 0.5–, lane 4, 0.25–, lane 3, 0.125–, lane 2, 0.0625–). A binding reaction with unprogrammed reticulocyte lysate was run in lane 7. Arrows on the right indicate the positions of the free DNA probe and the protein/DNA complexes. Arrows on the left indicate the position of the protein/DNA complex bands that represents the homo- and hetero-dimers.

Conclusions

Sequence similarity of EmBP-1 with GCN4 and other members of the bZIP family of transcription factors suggested that EmBP-1 shares features common to these proteins; ie. the basic DNA-binding domain, the leucine dimerization domain and the transcriptional activation domain [10]. However, prior to this work, these assumptions remained to be directly tested for EmBP-1. The results presented here demonstrate that EmBP-1 shares a high degree of similarity to GCN4 in the location and function of its basic and leucine zipper regions. We have shown that the entire conserved basic domain is necessary for DNA-binding and that deletion of three or more of the leucine repeat elements reduces or eliminates DNA-binding. Furthermore, as expected, EmBP-1 binds DNA as a dimer. The unique leucine-proline domain of EmBP-1 does not appear to be important in DNA-binding or dimerization, but this does not rule out the possibility of its importance in other functions such as transcriptional activation or interactions with accessory proteins. The high degree of functional conservation of the bZIP proteins spanning organisms from plants to fungi highlights the ancient origin of this class of transcription factors and of the mechanisms of gene regulation in which they participate.

Acknowledgements

We would like to thank Drs Jill Deikman, June Medford and Ross Hardison for critical review of this manuscript and Patricia Hughes for her 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. Agre P, Johnson PF, McKnight SL: Cognate DNA-binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. Science 246: 922–925 (1989).
2. Chen M, Middledorp JM, Hayward SD: Separation of the complex DNA-binding domain of EBNA-1 into DNA recognition and dimerization subdomains of novel structure. J Virol 67: 4875–4885 (1993).
3. Dash PK, Moore AN: A peptide containing the leucine zipper domain specifically inhibits CREB binding and transcription. Cell Mol Biol 39: 35–43 (1993).
4. DeVos KM, Atkinson MD, Chinoy CN, Guiltinan MG, Quatrano RS, Gale MD: Chromosomal location and variability in wheat, barley and rye of a wheat gene encoding a bZIP protein (EmBP-1). Theor Appl Genet 82: 665–667 (1991).
5. Donald RGK, Schindler AB, Cashmore AR: The plant G box promoter sequence activates transcription in *Saccharomyces cerevisiae* and is bound *in vitro* by a yeast

activity similar to GBF, the plant G box binding factor. EMBO J 9: 1727–1735 (1990).
6. Ehrlich KC, Cary JW, Ehrlich M: A broad bean cDNA clone encoding a DNA-binding protein resembling mammalian CREB in its sequence specificity and DNA methylation sensitivity. Gene 117: 169–178 (1992).
7. Ellenberger TE, Brandel CJ, Struhl K, Harrison SC: 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 (1992).
8. Foley RC, Grossman C, Ellis JG, Llewellyn DJ, Dennis ES, Peacock WJ, Singh KB: Isolation of a maize bZIP protein subfamily: candidates for the ocs-element transcription factor. Plant J 3: 669–679 (1993).
9. Green S: Promicuous liasons. Nature 361: 590–591 (1993).
10. Guiltinan MJ, Marcotte WR Jr., Quatrano RS: A plant leucine zipper protein that recognizes an abscisic acid response element. Science 250: 267–271 (1990).
11. Hope I, Struhl K: GCN4 protein, synthesized *in vitro*, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. Cell 43: 177–188 (1985).
12. Hope I, Struhl K: GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J 6: 2781–2784 (1987).
13. Hope IA, Struhl K: Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46: 885–894 (1986).
14. Izawa T, Foster R, Chua N-H: Plant bZIP protein binding specificity. J Mol Biol 230: 1131–1144 (1993).
15. Jones N: Transcriptional regulation by dimerization: two sides to an incestuous relationship. Cell 61: 9–11 (1990).
16. Katagiri F, Lam E, Chua N-H: Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature 340: 727–730 (1989).
17. Kawata T, Imada T, Shiraishi H, Okada K, Shimura Y, Iwabuchi M: A cDNA clone encoding HBP-1b homologue in *Arabidopsis thaliana*. Nucl Acids Res 20: 1141 (1992).
18. Klug: Co-chairman's remarks: protein designs for the specific recognition of DNA. Gene 135: 83–92 (1993).
19. Marcotte WR, Bayley CC, Quatrano RS: Regulation of a wheat promoter by abscisic acid in rice protoplasts. Nature 335: 454–457 (1988).
20. Marcotte WR, Russel SH, Quatrano RS: Abscisic acid-responsive sequences from the Em gene of wheat. Plant Cell 1: 969–979 (1989).
21. Mundy J, Chua N-H: Abscisic acid and water-stress induce the expression of a novel rice gene. EMBO J 7: 2279–2286 (1988).
22. Nagaya T, Jameson JL: Thyroid hormone receptor dimerization is required for dominant negative inhibition by mutations that cause thyroid hormone resistance. Biol Chem 268: 15766–15771 (1993).
23. O'Shea EK, Klemm JD, Kim PS, Alber T: X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 254: 539–544 (1991).
24. O'Shea EK, Rutkowski R, Stafford WF, Kim PS: Preferential heterodimer formation by isolated zippers from Fos and Jun. Science 245: 646–648 (1989).
25. Oeda K, Salinas J, Chua NH: A tobacco bZIP transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J 10: 1793–1991 (1991).
26. Pla M, Vilardell J, Guiltinan MJ, Marcotte WR, Niogret MF, Quatrano RS, Pages M: The cis-regulatory element CCACGTGG is involved in ABA and water-stress responses of the maize gene *rab28*. Plant Mol Biol 21: 259–266 (1993).
27. Prendergast GC, Ziff EB: Methylation-sensitive sequence-specific DNA-binding by the c-Myc basic region. Science 251: 186–189 (1991).
28. Pysh LD, Aukerman MJ, Schmidt RJ: OHP1: a maize basic domain/leucine zipper protein that interacts with opaque2. Plant Cell 5: 227–236 (1993).
29. Quatrano RS, Guiltinan MJ, Marcotte WR: Regulation of gene expression by abscisic acid. In: Verma DPS (ed) The Control of Gene Expression, pp. 69–90. CRC Press, Boca Raton, FL (1992).
30. Schagger H, von Jagow G: Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the seperation of proteins in the range from 1 to 100 kDa. Anal Biochem 166: 368–379 (1987).
31. Schindler J, Beckmann H, Cashmore AR: 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 (1992).
32. Schindler U, Menkens AE, Beckmann H, Ecker JR, Cashmore AR: Heterodimerization between light-regulated and ubiquitously expressed *Arabidopsis* GBF bZIP proteins. EMBO J 11: 1261–1273 (1992).
33. Schindler U, Terzaghi W, Beckmann H, Kadesch T, Cashmore AR: DNA-binding site preferences and transcription activation properties of the *Arabidopsis* transcription factor GBF-1. EMBO J 11: 1275–1289 (1992).
34. Schmidt RJ, Burr FA, Aukerman MJ, Burr B: Maize regulatory gene opaque-2 encodes a protein with a leucine-zipper motif that binds to zein DNA. Proc Nat Acad Sci USA 87: 46–50 (1990).
35. Singh K, Dennis ES, Ellis JG, Llewellyn DJ, Tokuhisa JG, Wahleither JA, Peacock WJ: OCSBF-1, a maize ocs enhancer binding factor: isolation and expression during development. Plant Cell 2: 891–903 (1990).
36. Skriver K, Lok Olsen F, Rogers JC, Mundy J: Cis-acting DNA elements responsive to gibberellin and its antagonist abscisic acid. Proc Nat Acad Sci USA 88: 7266–7270 (1991).
37. Tabata T, Nakayama T, Mikami K, Iwabuchi M: HBP-1a and HBP-1b: leucine zipper-type transcription factors of wheat. EMBO J 10: 1459–1467 (1991).
38. Tabata T, Takase H, Takayama S, Mikami K, Nakatsuka A, Kawata T, Nakayama T, Iwabuchi M: A protein

that binds to a cis-acting elemnet of wheat histone genes has a leucine zipper motif. Science 245: 965–967 (1989).
39. Vilardell J, Goday AJ, Freire MA, Torrent M, Martinez MC, Torne JM, Pages M: Gene sequence, developmental expression and protein phosphorylation of RAB-17 in maize. Plant Mol Biol 14: 423–432 (1990).
40. Vinson CR, Sigler PB, McKnoght SL: Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science 246: 911–916 (1989).
41. Voronova A, Baltimore D: Mutations that disrupt DNA-binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proc Natl Acad Sci USA 87: 4722–4726 (1990).
42. Watt F, Molloy PL: Specific cleavage of transcription factors by the thiol protease, m-calpain. Nucl Acids Res 21: 5092–5100 (1993).
43. Weisshaar B, Armstrong GA, Block A, Silva OC, Hahlbrock K: Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. EMBO J 10: 1777–1786 (1991).
44. Williams ME, Foster R, Chua N-H: Sequences flanking the hexameric G-box core CACGTG affect the specifcity of protein binding. Plant Cell 4: 485–496 (1992).
45. Xue D, Tu Y, Chalfie M: Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3. Science 261: 1324–1328 (1993).
46. Zhang B, Foly RC, Singh KB: Isolation and characterization of two related *Arabidopsis ocs*-element bZIP binding proteins. Plant J. 4: 711–716 (1993).
47. Zhang X, Pfahl M: Regulation of retinoid and thyroid hormone action through homodimeric and heterodimeric receptors. Trends Endo Metab 4: 156–162 (1993).