The EMBO Journal vol.11 no.4 pp.1261 - 1273, 1992

Heterodimerization between light-regulated and ubiquitously expressed *Arabidopsis* GBF bZIP proteins

Ulrike Schindler, Anne E.Menkens, Holger Beckmann, Joseph R.Ecker and Anthony R.Cashmore

University of Pennsylvania, Plant Science Institute, Department of Biology, Philadelphia, PA 19104-6018, USA

Communicated by M.van Montagu

The promoters of a variety of plant genes are characterized by the presence of a G-box (CCACGTGG) or closely related DNA motifs. These genes often exhibit quite diverse expression characteristics and in many cases the G-box sequence has been demonstrated to be essential for expression. The G-box of the *Arabidopsis rbcS-1A* gene is bound by a protein, GBF, identified in plant nuclear extracts. Here we report the isolation of three *Arabidopsis thaliana* cDNA clones encoding GBF proteins referred to as GBF1, GBF2 and GBF3. *GBF1* and *GBF2* mRNA is present in light and dark grown leaves as well as in roots. In contrast, *GBF3* mRNA is found mainly in dark grown leaves and in roots. The deduced amino acid sequences of the three cDNAs indicate that each encodes a basic/leucine zipper protein. In addition, all three proteins are characterized by an N-terminal proline-rich domain. Homodimers of the three proteins specifically recognize the G-box motif, with GBF1 and GBF3 binding symmetrically to this palindromic sequence. In contrast, GBF2 binds to the symmetrical G-box sequence in such a way that the juxtaposition of the protein and the DNA element is clearly asymmetric and hence distinct from that observed for the other two proteins. The fact that GBF1, GBF2 and GBF3 possess both distinct DNA binding properties and expression characteristics prompt us to entertain the notion that these proteins may individually mediate distinct subclasses of expression properties assigned to the G-box. Furthermore, we demonstrate that GBF1, GBF2 and GBF3 heterodimerize and these heterodimers also interact with the G-box, suggesting a potential mechanism for generating additional diversity from these GBF proteins.

Key words: Arabidopsis bZIP proteins/GBF proteins/heterodimerization/light regulation/proline-rich region

Introduction

Regulation of gene expression is directed in part by the action of specific transcription factors, which in response to external and internal stimuli interact with three cognate DNA binding sites. Previously we have shown that the promoters for RBCS genes in plants are often characterized by conserved DNA sequences (Giuliano *et al.*, 1988). One of these motifs, the G-box, is defined by the palindromic sequence CCACGTGG. In the case of the *Arabidopsis thaliana rbcS-1A* promoter, we demonstrated that this motif is required for expression of a GUS reporter gene in both transgenic tobacco and *Arabidopsis* plants (Donald and Cashmore, 1990). *In vitro* DNA binding studies revealed that the G-box is specifically bound by the nuclear factor GBF (Giuliano *et al.*, 1988).

G-boxes and DNA motifs differing by 1 or 2 bp substitutions (hereafter referred to as G-box-like elements) are also found in other plant promoters. Some of these promoters, including the *cab-E* promoter of *Nicotiana plumbaginifolia* and the chalcone synthase (*chs*) promoters of *Petroselinum crispum* and *Antirrhinum majus* activate transcription upon light induction. Others respond to anaerobiosis, such as the *Arabidopsis* alcohol dehydrogenase (*adh*) promoter, or to hormones like ABA. Expression studies have demonstrated that the activity of these different promoters requires the G-box or G-box-like motifs (Block *et al.*, 1990; Castresana *et al.*, 1988; Marcotte *et al.*, 1989; Mundy *et al.*, 1990; Schulze-Lefert *et al.*, 1989). Moreover, *in vitro* and *in vivo* binding studies revealed that these elements are bound by GBF-like nuclear proteins (DeLisle and Ferl, 1990; Ferl and Laughner, 1989; McKendree *et al.*, 1990; Mundy *et al.*, 1990; Schindler and Cashmore, 1990; Schulze-Lefert *et al.*, 1990; Staiger *et al.*, 1989, 1991).

Recently, cDNAs encoding nuclear proteins that interact with G-box and G-box-like promoter sequences were isolated from different plant species. The wheat protein EmBP-1, for example interacts with a sequence motif that is part of the ABA responsive element of the wheat *Em* gene (Guiltinan *et al.*, 1990). The three parsley proteins CPRF-1, CPRF-2 and CPRF-3 bind to box II of the parsley *chs* promoter. Box II, which resembles the G-box, resides within a promoter fragment that is required for the UV induction of the *chs* gene (Weisshaar *et al.*, 1991). The tobacco protein TAF-1 interacts with motif I, a G-box-like element found in the ABA induced rice *rab* promoters (Oeda *et al.*, 1991). Recently it was demonstrated that the wheat protein HBP-1a, previously designated HBP-1, also recognizes a G-box motif (Tabata *et al.*, 1989, 1991). All these proteins belong to the class of basic/leucine zipper (bZIP) proteins.

The class of bZIP proteins is characterized by their bipartite DNA binding domain that consists of a region enriched in basic amino acids (basic region, BR) with an adjacent leucine zipper (Landschulz *et al.*, 1988; Vinson *et al.*, 1989). The characteristic features of the leucine zipper are the periodic repetition of leucines at every seventh position, interdigitated with other hydrophobic residues at every fourth position (O'Shea *et al.*, 1989). This region forms an amphipathic α-helix and the helices of two polypeptides, oriented in a parallel manner, are thought to associate in a coiled coil configuration resulting in dimerization (Hu *et al.*, 1990; Rasmussen *et al.*, 1991; and references therein).

bZIP proteins commonly occur in families whose members bind to the same or related DNA motifs. Some of the best characterized families include the mammalian proteins Fos

and Jun (Ziff, 1990 and references therein), CREB/ATF (Gonzalez *et al.*, 1989; Hai *et al.*, 1989; Hai and Curran, 1991; Hoeffler *et al.*, 1988) and C/EBP (Cao *et al.*, 1991; Williams *et al.*, 1991). In addition to interacting with DNA as the homodimer, members of these families have the capacity to form heterodimers. These heterodimers form *in vivo* and are often characterized by distinct binding properties (Hai and Curran, 1991; Williams *et al.*, 1991). This capacity to heterodimerize clearly increases the potential number of transcription factors available for any regulatory system.

We are interested in exploring the mechanism by which the G-box and G-box-like elements mediate a diverse array of expression characteristics. An understanding of the potential regulatory information inherent in these DNA motifs requires identification of the entire complement of proteins that recognize these DNA sequences and knowledge of the tissues in which each member is present and active. Detailed analysis of the DNA binding properties of these proteins is important from the viewpoint that any differences could be utilized in distinguishing between these proteins. As a start towards addressing these questions we isolated three *Arabidopsis* cDNAs encoding proteins that interact with the G-box motif. The three proteins, designated GBF1, GBF2 and GBF3, heterodimerized with each other and these heterodimers also interacted with the G-box. *GBF3* mRNA was detected mainly in roots and dark grown leaves and very little in light-grown tissue; in contrast, at this level of analysis, the expression of *GBF1* and *GBF2* appeared to be constitutive. These results enhance our understanding of how the G-box can mediate a diverse array of expression characteristics.

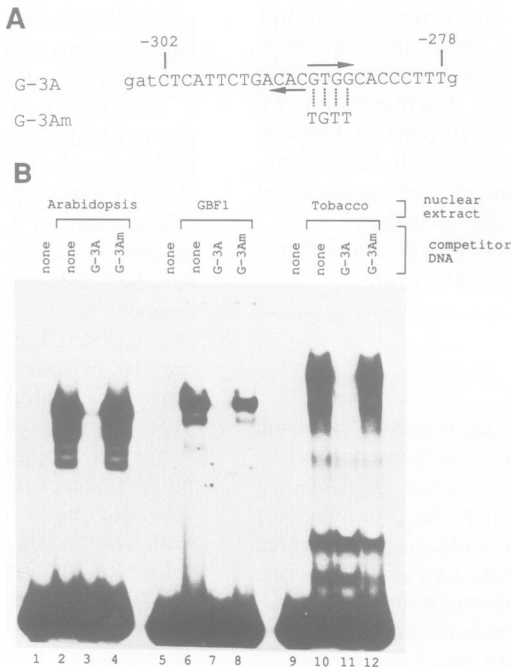
Results

Isolation of Arabidopsis cDNAs encoding GBF1

In order to identify cDNAs encoding G-box binding proteins, we screened an *Arabidopsis* cDNA expression library for DNA binding activity specific to a synthetic oligonucleotide bearing the tomato *rbcS-3A* G-box-like element (G-3A, Figure 1A). Initial DNA binding studies employing *Arabidopsis* nuclear extracts had demonstrated that this DNA element was bound with an affinity similar to that of the palindromic *Arabidopsis rbcS-1A* G-box element (Donald *et al.*, 1990; Giuliano *et al.*, 1988). Two positive phage were detected in the initial screen and subsequently purified to homogeneity. The cDNAs, *GBF1* and *GBF1a*, were subcloned and propagated in *Escherichia coli*. The DNA binding specificities and the relative molecular masses of the proteins encoded by these two cDNAs were determined by South-western analysis using the wild-type G-3A oligonucleotide and a mutant derivative (G-3Am; Figure 1A). Although the recombinant proteins differed in their relative molecular mass (39 and 32 kDa), both interacted with the wild-type G-3A oligonucleotide but not with the mutant G-3Am. No binding was obtained with bacterial extracts not expressing GBF1 or GBF1a (Schindler *et al.*, 1991). These results confirmed that the isolated cDNAs encode proteins that specifically recognize the G-box-like sequence of the tomato *rbcS-3A* promoter.

GBF1 exhibits DNA binding properties similar to GBF found in crude nuclear extracts

We used competitive DNA binding assays to compare the DNA binding properties of GBF1 with the GBF activity present in *Arabidopsis* and tobacco nuclear extracts. Tobacco extracts were used for these experiments since we have previously demonstrated the essential role of the G-box in *rbcS-1A* promoter activity in transgenic tobacco plants. The DNA probe used for the binding studies spanned the *rbcS-1A* promoter from position −251 to −211. As shown in Figure 1B, in addition to the major DNA−protein complex obtained with the nuclear extracts, several more slowly migrating complexes were also observed (lanes 2 and 10). All of these complexes were specifically competed when the G-3A oligonucleotide of the tomato *rbcS-3A* promoter was included in the binding reaction (lanes 3 and 11). No competition was observed with the mutant G-3Am oligo-



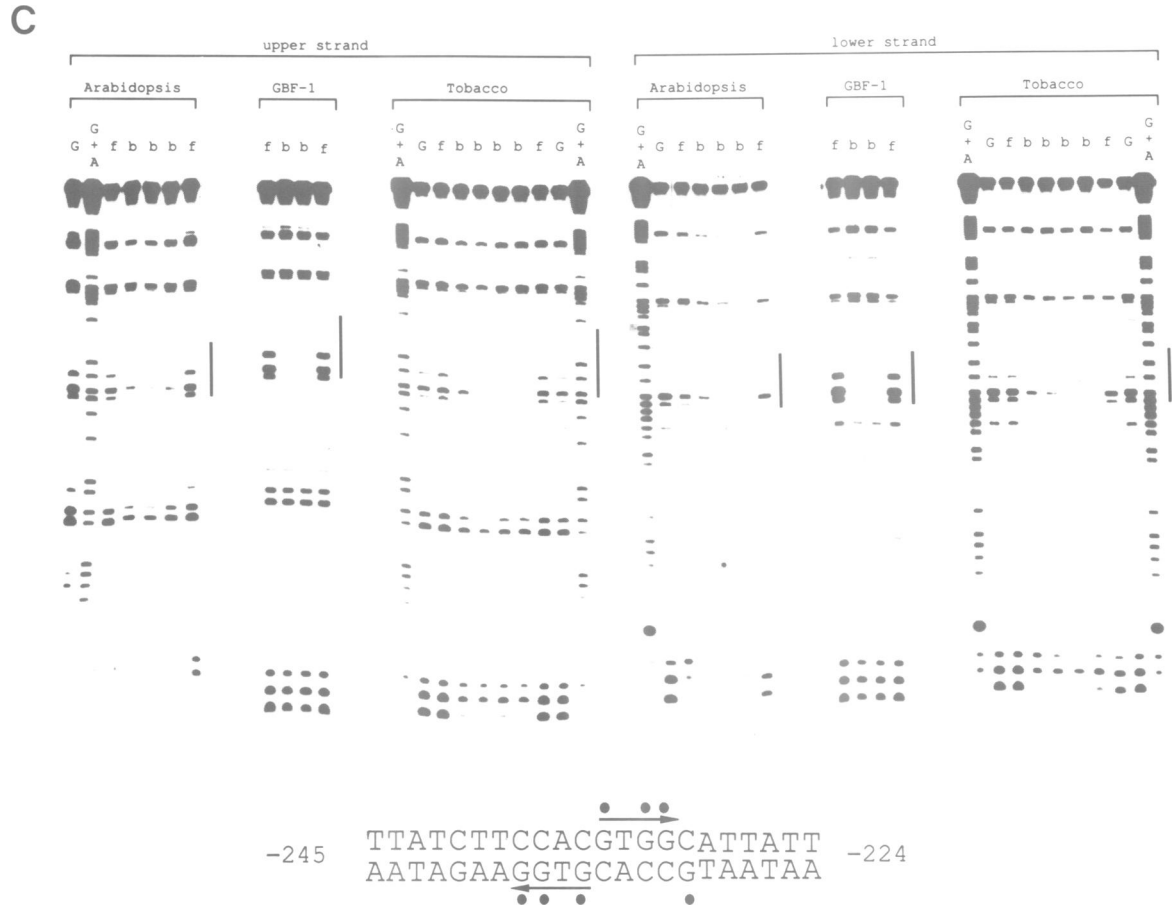


Fig. 1. GBF1 and GBF, isolated from crude nuclear extract, show the same DNA binding specificity. (**A**) Synthetic oligonucleotides used to determine the binding specificity. G-3A represents the tomato *rbcS-3A* G-box-like element; G-3Am, a mutant derivative, carries the indicated 4 bp substitutions. The location of the G-box is marked by arrows. Numbers correspond to the nucleotide positions within the tomato *rbcS-3A* promoter. (**B**) Competition assay using either recombinant GBF1 or crude plant nuclear extracts. The G-box-containing *rbcS-1A* promoter fragment (−251 to −211) was incubated with either no extracts (lanes 1, 5 and 9), 10 μg *Arabidopsis* crude nuclear extracts (lanes 2−4), 0.5 μg bacterial extracts containing recombinant GBF1 (lanes 6−8) or 10 μg tobacco extracts (lanes 10−12). The binding reactions were performed in the absence of specific competitor DNA (lanes 2, 6 and 10) or in the presence of 1 pmol G-3A (lanes 3, 7 and 11) or 10 pmol G-3Am oligonucleotides (lanes 4, 8 and 12). (**C**) Methylation interference assays employing the same promoter fragment and proteins (80 μg plant nuclear extract or 2.5 μg bacterial extracts) as shown in Figure 1B. The free (f) and protein-complexed (b) forms of both partially methylated DNA strands (upper and lower) were isolated separately and after piperidine cleavage analyzed on an 8% denaturing polyacrylamide gel. In all six panels, the first 'bound' band corresponds to the most rapidly migrating protein−DNA complex illustrated in B. G and G+A represent marker lanes derived from the same DNA fragment subjected to Maxam−Gilbert sequencing reactions (Maxam and Gilbert, 1980). The bars mark the position of the G-box. A summary of the results is shown below. Closed circles indicate the G residues which, when methylated, interfere with protein binding, the arrows designate the palindromic nature of the element. Numbers correspond to the nucleotide positions within the *Arabidopsis rbcS-1A* promoter.

nucleotide (lanes 4 and 12). It is not known if the different protein−DNA complexes were due to proteolytic breakdown products of a single polypeptide (or some other post-translational modification) or if they were due to multiple GBF-like DNA binding proteins. The recombinant protein, GBF1, formed a complex with the *rbcS-1A* promoter fragment with an electrophoretic mobility similar to that of the major complex formed with the *Arabidopsis* nuclear extracts (lane 6). The DNA binding specificity of GBF1, as determined by competitive binding studies (lanes 7 and 8), was the same as that observed for the activity in the crude nuclear extracts. The slight competition obtained with the G-3Am oligonucleotide (lane 8) may reflect the high amount of competitor DNA used relative to the bacterial protein.

To explore further the similarities between GBF1 and the G-box binding activity found in crude nuclear extracts, we performed methylation interference studies. Figure 1C demonstrates that a similar methylation interference pattern was obtained with recombinant GBF1 and *Arabidopsis* or tobacco nuclear extracts. Methylation of any of the six G residues within the core recognition sequence CCACGTGG prevented protein binding. The multiple protein−DNA complexes obtained in the mobility shift assay (Figure 1B) were analyzed separately; all of them centered over the same DNA binding site, the G-box (Figure 1C). These studies demonstrated that GBF1 exhibited DNA binding properties similar to the G-box binding protein(s) present in the *Arabidopsis* nuclear extract.

GBF1 is a basic/leucine zipper protein with an N-terminal proline-rich region

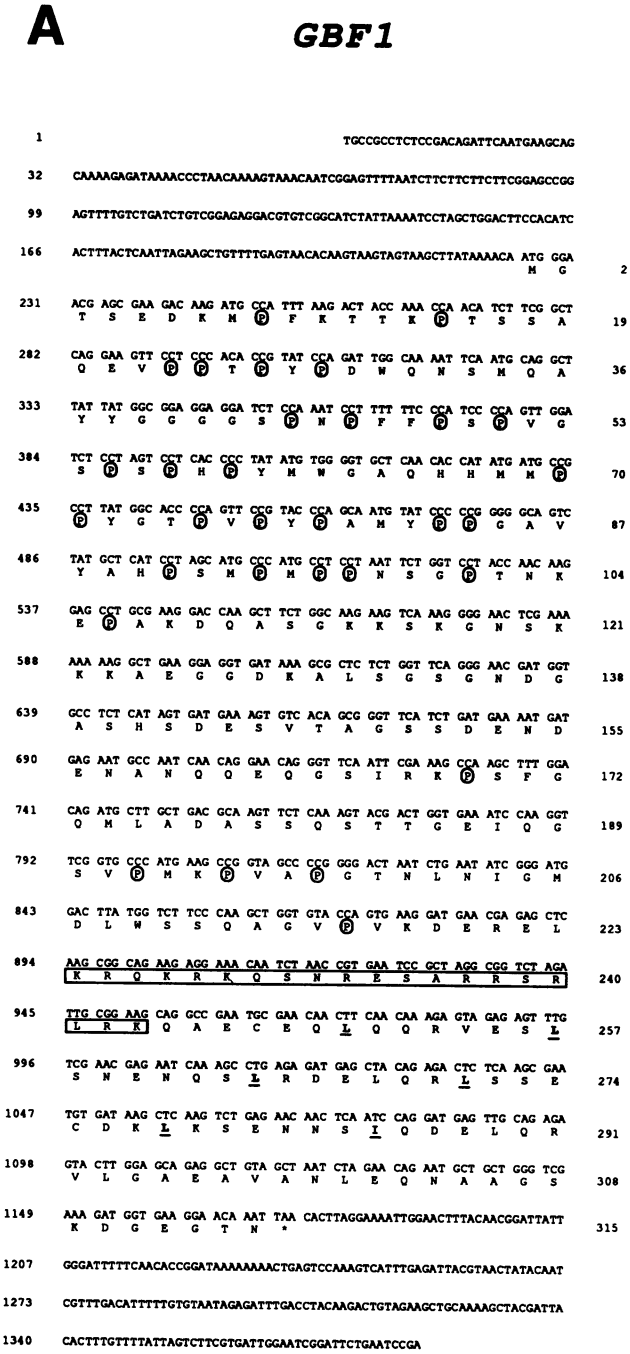
Sequence analysis of both cDNAs, *GBF1* and *GBF1a*, indicated that they were derived from the same mRNA species. The DNA and the deduced amino acid sequence of the longer of the two cDNAs (*GBF1*) is shown in Figure 2A. The longest open reading frame started with an AUG codon at position 225 and encoded a 34 kDa protein. The shorter cDNA, *GBF1a*, corresponded to bp 570−1388

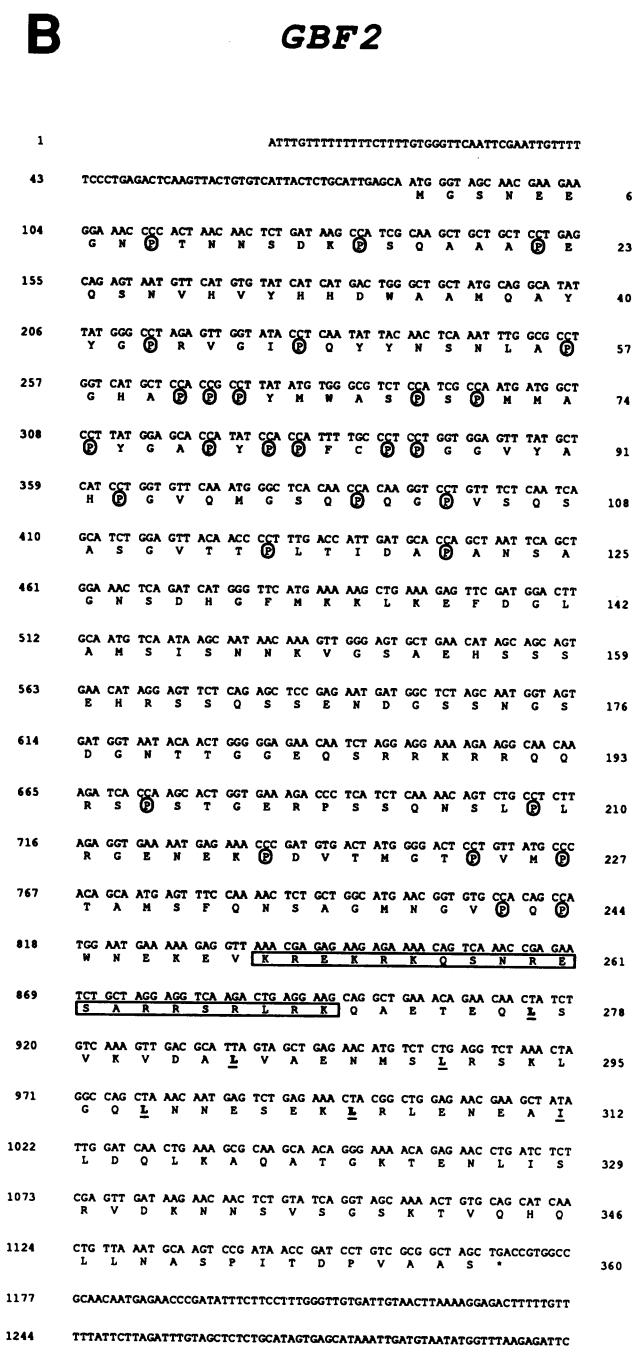
and encoded a 21 kDa protein. In contrast to the 34 kDa protein, the 21 kDa protein was fused in-frame to the N-terminal domain of β-galactosidase.

The analysis of the GBF1 amino acid sequence revealed two interesting regions. The first, located in the N-terminus of the protein, is characterized by its high proline content (24.5% proline within the first 106 amino acids, Figure 2A). This region functions as an activating domain when fused to a heterologous DNA binding domain (Schindler *et al.*, 1992). The second region of GBF1, encompassing amino acids 250–285, is a leucine zipper domain consisting of the five heptamer repeats with every seventh position occupied by a leucine or isoleucine residue. The leucine zipper is located next to a region rich in basic amino acids (amino acids 224–243). Hence, GBF1 belongs to the class of bZIP proteins (Landschulz *et al.*, 1988).

Isolation of two other Arabidopsis cDNAs encoding GBF2 and GBF3

The G-box is crucial for the expression of a variety of plant genes, many of which are induced by different stimuli. In view of this we were interested in investigating how many G-box binding proteins existed in a simple diploid plant such as *Arabidopsis*. To isolate other cDNA clones, a DNA probe corresponding to the DNA binding domain of GBF1 was used to screen our *Arabidopsis* cDNA library by DNA hybridization. Two positively hybridizing phage were isolated and shown to contain cDNA inserts of 1.4 kb and





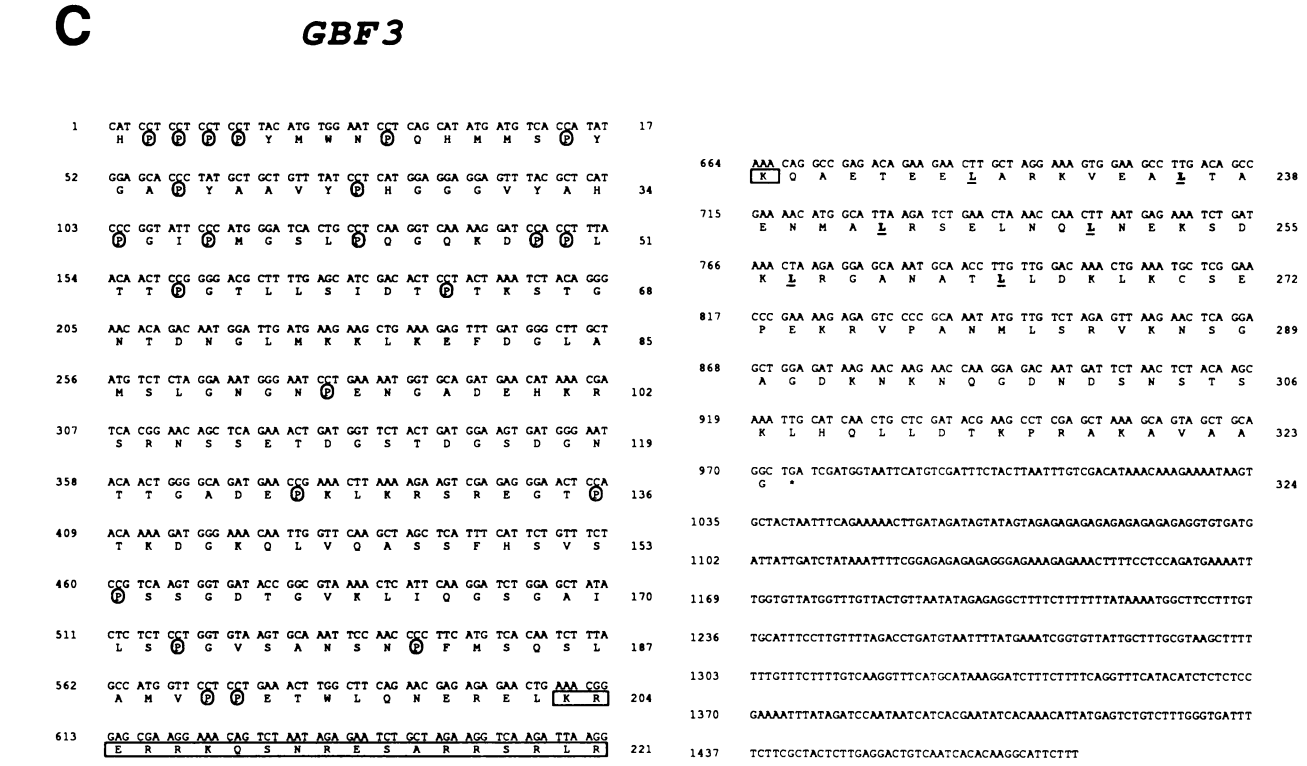


Fig. 2. DNA sequence of the cDNAs encoding GBF1, GBF2 and GBF3 and predicted amino acid sequence. The DNA sequence and deduced amino acid sequence of *GBF1* (**A**), *GBF2* (**B**) and *GBF3* (**C**) are shown. Numbers on the left and right correspond to the base pair and amino acid positions respectively. The proline residues within the N-terminal domains are circles. The amino acids representing the basic region are boxed, the leucine residues within the leucine zipper domains are in bold and underlined.

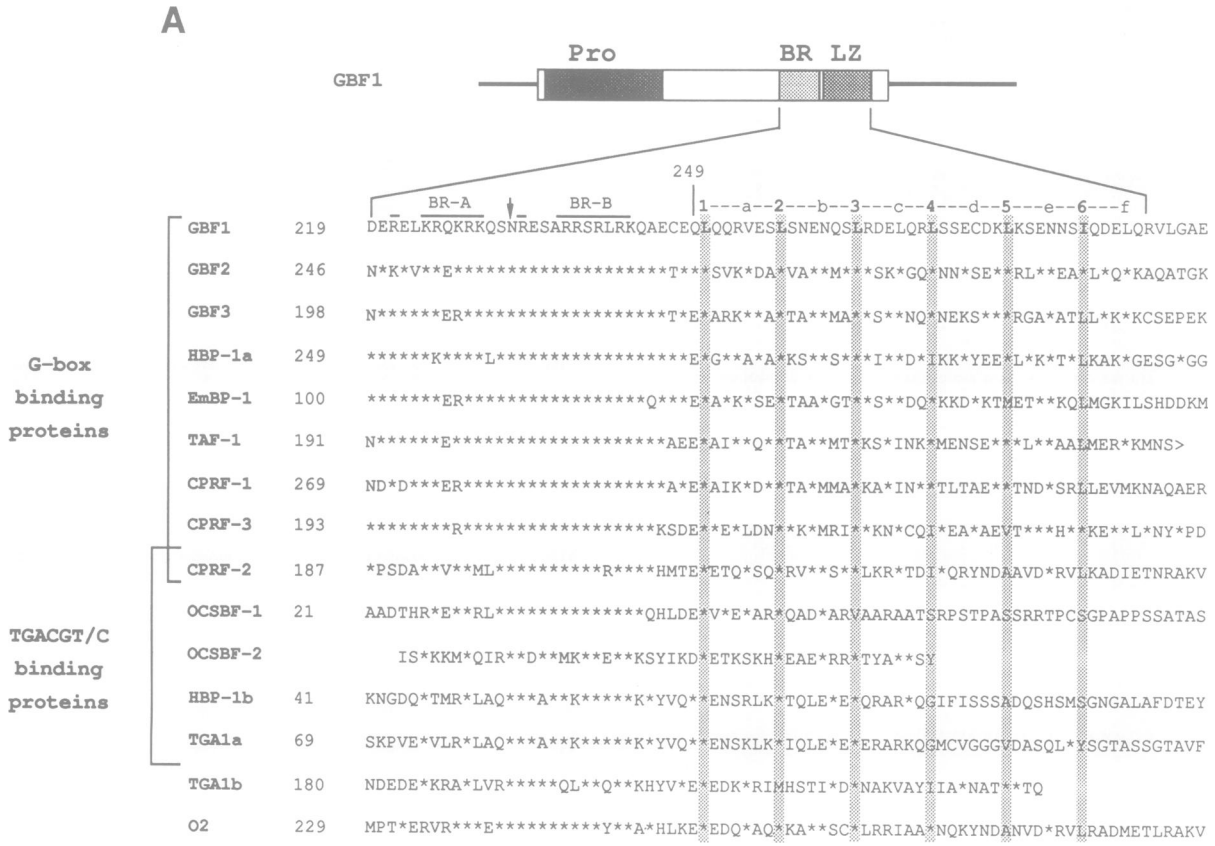
1.5 kb. DNA sequencing analysis demonstrated that these cDNAs, designated *GBF2* and *GBF3* respectively, also encoded bZIP proteins with N-terminal proline-rich regions (Figure 2B and C). The cDNA encoding GBF3 is probably not full-length since the open reading frame starts with an asparagine which we have tentatively designated as the first amino acid. For reasons we do not understand, none of the three cDNAs contain a poly(A) tail.

The putative DNA binding domains of GBF2 and GBF3 are very similar to the one identified in GBF1 (Figure 3A). Within the basic region (20 amino acids) these two proteins have 95% and 90% amino acid identity with GBF1, respectively. The basic regions of all three proteins consist of two clusters of basic amino acids, designated basic region A (BR-A) and B (BR-B, Figure 3A). These two clusters are separated by an asparagine found in all bZIP proteins identified so far (McKnight, 1991; Pu and Struhl, 1991). Both BR-A and BR-B in the three *Arabidopsis* GBF proteins are very similar to the corresponding regions in other plant bZIP proteins which have been shown to interact with G-box-like promoter elements (Figure 3A). In contrast, when the DNA binding domains of GBF1, GBF2 and GBF3 were compared with the bZIP proteins TGA1a, TGA1b (Katagiri *et al.*, 1989), OCSBF-1, OCSBF-2 (Singh *et al.*, 1990) and HBP-1b (Tabata *et al.*, 1991), all of which have been shown to interact with the TGACGT/C motifs identified in viral and plant promoters, a strong sequence similarity was observed only within BR-B (Figure 3A). Furthermore, the parsley protein CPRF-2, which interacts with a G-box and the TGACGT/C motifs of the CaMV 35S promoter (as-1 element) is also only similar to the *Arabidopsis* GBFs within BR-B (Weisshaar *et al.*, 1991). The same observation was made for the maize protein Opaque-2 which recognizes distinct but related DNA motifs (Lohmer *et al.*, 1991).

The N-terminal domains of GBF1, GBF2 and GBF3 are rich in proline residues. Although GBF2 and GBF3 are very similar within this region (55 out of 92 amino acids), the two proteins are more divergent from GBF1 (27 out of 68 and 20 out of 45 amino acids respectively, Figure 3B). Interestingly, GBF1 is actually more similar to the wheat protein HBP-1a (34 out of 64 amino acids) and the parsley protein CPRF-3 (39 out of 64 amino acids) than it is to GBF2 and GBF3, indicating a strong evolutionary conservation and implying some related regulatory function. The subtle differences between GBF1, GBF2 and GBF3 within these domains may reflect the possibility that these proteins are involved in distinct signal transduction pathways.

All three bZIP proteins interact with the G-box but not with the TGACGT/C motifs of the CaMV 35S promoter

The strong sequence similarity between GBF1, GBF2 and GBF3 within their DNA binding domains suggested that the two latter proteins—like GBF1—should interact with G-box and G-box-like elements. To confirm this hypothesis, competitive DNA binding studies employing all three proteins were performed. The proteins were expressed *in vitro* using a rabbit reticulocyte lysate translation system. The translation products were incubated with a radiolabeled oligonucleotide bearing the G-box element of the *Arabidopsis rbcS-1A* promoter (G-1A, Figure 4A). The *in vitro* translation products of the partial cDNA *GBF3* bound very inefficiently to the G-1A oligonucleotide. This inefficient DNA binding may be due to inappropriate folding of the



B

```
GBF3                          1 HPPPPYMWN-PQHMMS-PYGAPYAAVYPHG--GGVYAHPGIPMGSLPQGQKDPPL---TTPGTLLSIDTPTKSTGNTDNGLMKKLKEFDGLAMSLGNGN 92
GBF2     36 AMQAYYGPRVGIPQ-YYNSNLAPGHAPPPYMWASPSPMMA-PYGAPYPPFCPP---GGVYAHPGVQMGSQPQGPVSQSASGVTTPLTIDAPANSA---GNSDHGFMKKLKEFDGLAMSISNNK 150
CPRF-1   35 AMQAYYGPRVALPP-YFNPAVASGQSPHPYMWGPPQPVMP-PYGVPYAALYAH---GGVYAHPGVPLAASPMSMDTHAKSSGTNEH----------------GLIKKLKGHDDLAMSIGNGK 135
CPRF-3   35 SMQAYYGGAA-PAAFYASTVGSPS--PHPYMWRN-QHRFILPYGIPMQ-YPALFLPGGIFTHPIVPTDPNLAP 102
GBF1     33 SMQAYYGGGGSPNPFFPSPVGSPS--PHPYMWGA-QHHMMPPYGTP-VPYPAMYPPGAVYAHPSMPMPPNSGP 101
HBP-1a   36 WPGFQGYPAMPPHGFFPPPVAAGQA--HPYMWG-PQH-MVPPYGTPPPPY-MMYPPGTVYAHPTAPGVHPFHY 103
```

Fig. 3. Similarity between GBF1, GBF2, GBF3 and other plant bZIP proteins. (**A**) Top: schematic presentation of GBF1, indicating the location of the proline-rich region (Pro), the basic region (BR) and the leucine zipper domain (LZ). Bottom: amino acid comparison of the DNA binding domain of GBF1 and other plant bZIP proteins (Guiltinan *et al.*, 1990; Katagiri *et al.*, 1989; Oeda *et al.*, 1991; Schmidt *et al.*, 1990; Singh *et al.*, 1990; Tabata *et al.*, 1989, 1991; Weisshaar *et al.*, 1991). Proteins binding to the G-box or G-box-like elements of various plant promoters and proteins interacting with the TGACGT/C motifs of either the CaMV 35S promoter or the nopaline synthase or octopine synthase promoters were classified as G-box or TGACGT/C binding proteins. Numbers on the left refer to the amino acid positions within the individual proteins. Amino acids identical to GBF1 are indicated by asterisks. BR-A and BR-B designate the two clusters of basic amino acids (overlined), which are separated by the asparagine (arrow) found in all bZIP proteins identified so far (McKnight, 1991; Pu and Struhl, 1991). The leucine residues within the leucine zipper are highlighted and numbered. The hydrophobic amino acids interdigitating the leucine repeats are marked a−f. (**B**) Amino acid similarities within the N-terminal proline-rich regions of GBF1, GBF2, GBF3, the two parsley proteins CPRF-1 and CPRF-3 (Weisshaar *et al.*, 1991) and the wheat protein HBP-1a (Tabata *et al.*, 1989). Vertical bars indicate identical amino acids. Numbers refer to the amino acid positions within the individual proteins. Gaps (−) were introduced to optimize the alignment.

truncated GBF3 protein *in vitro*. Therefore a shorter version of GBF3 (amino acids 95−324) was employed in the following studies. The results illustrated in Figure 4B demonstrated that all three proteins GBF1, GBF2 and the truncated derivative of GBF3, interacted with the G-1A oligonucleotide (lanes 3, 11 and 19). The protein−DNA complexes were specifically competed when either the homologous unlabeled oligonucleotide G-1A (lanes 4, 12 and 20) or an oligonucleotide containing the G-box-like element of the tomato *rbcS-3A* promoter (G-3A, Figure 1A) were included in the binding reactions (Figure 4B, lanes 5, 13 and 21). In contrast, no competition was obtained with the mutant derivative G-3Am (Figures 1A and 4B, lanes 6, 14 and 22). Furthermore, all three proteins interacted with an oligonucleotide carrying the G-box element and surrounding sequences of the *Arabidopsis adh* promoter (Figure 4A and B, lanes 7, 15 and 23). Parts of the basic regions of GBF1, GBF2 and GBF3 are similar to other plant bZIP proteins which were shown to bind to the TGACGT/C motifs of the CaMV 35S promoter (as-1 element). In order to determine if either GBF1, GBF2 or GBF3 would interact with this latter element, an oligonucleotide bearing the as-1 element was designed (Figure 4A). Figure 4B illustrates that none of the three proteins interacted with the as-1 oligonucleotide (lanes 8, 16 and 24). These data conclusively demonstrate that all three GBF proteins specifically bind to G-box and G-box-

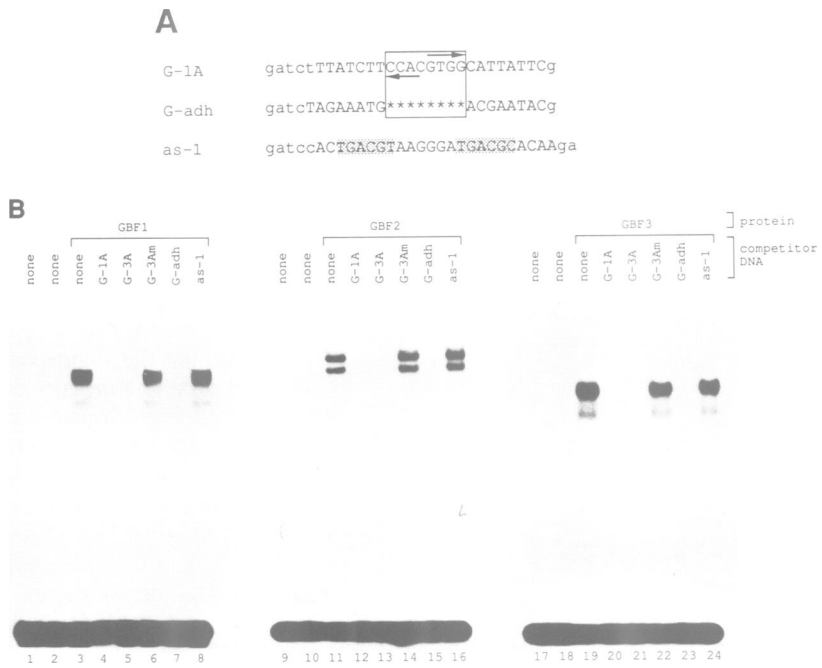


Fig. 4. GBF1, GBF2 and GBF3 interact with the G-box and G-box-like elements. (**A**) DNA sequence of the oligonucleotides employed in DNA binding studies. G-1A is derived from the *Arabidopsis rbcS-1A* promoter, G-adh from the *Arabidopsis adh* promoter and as-1 carries the two TGACGT/C motifs of the CaMV 35S promoter. (**B**) Competitive DNA binding studies employing GBF1, GBF2 and GBF3. The *in vitro* generated proteins GBF1 (lanes 3–8), GBF2 (lanes 11–16) and GBF3 (lanes 19–24) proteins were incubated with the radiolabeled G-1A oligonucleotide in the absence (lanes 3, 11 and 19) or presence (lanes 4–8, 12–16 and 20–24) of various competitor DNAs (1 pmol each) as indicated. The DNA sequence of these oligonucleotides is represented in panel A and Figure 1A. The individual translation products extend from amino acid 1 to 315 (GBF1), amino acids 1–360 (GBF2) and amino acids 95–324 (GBF3). Lanes 1, 9 and 17 contain the free DNA; lanes 2, 10 and 18 represent the binding reactions containing rabbit reticulocyte lysate only.

like motifs but not to the TGACGT/C motifs of the as-1 element.

GBF2 differs from GBF1 and GBF3 in that it binds in an asymmetric way to the symmetric G-box sequence within the Arabidopsis rbcS-1A promoter

As a step towards addressing the question of whether GBF1, GBF2 and GBF3 exhibit differences in their DNA binding characteristics, we performed methylation interference experiments employing the *Arabidopsis rbcS-1A* promoter fragment. The results obtained with all three proteins are illustrated in Figure 5. Both GBF1 and GBF3 bind in a symmetrical manner to the G-box and methylation of any one of the six G residues resulted in inhibition of binding. In contrast, binding to GBF2 was not inhibited when one of the center G residues at position −234 was methylated. Furthermore, methylation of the G residue at position −237 only partially prevented GBF2 binding. These results clearly demonstrate that binding of GBF2 involves a juxtaposition to the major groove of the DNA that is distinct from that found for GBF1 and GBF3. In addition, as the effect of methylation on the G residue at −234 is clearly distinct from the methylation of the G residue at −235—these G residues being symmetrically placed relative to the center of symmetry of the G-box (see sequence in Figure 5)—we conclude that GBF2 binds asymmetrically to the symmetrical and perfectly palindromic G-box of the *Arabidopsis rbcS-1A* promoter.

GBF1, GBF2 and GBF3 heterodimerize promiscuously

Leucine zipper proteins have been shown to interact with their cognate recognition sites as dimers (Vinson *et al.*, 1989). This dimer formation is mediated by the leucine zipper domain and is required for DNA binding (Talanian *et al.*, 1990). We investigated the ability of GBF1 to form dimers in the presence of DNA. Different portions of the cDNA encoding GBF1 were transcribed and translated *in vitro* and the protein products were employed in DNA binding studies. As shown in Figure 6A and B, when the full-length protein (GBF1L) and the truncated version (GBF1S) were individually assayed for DNA binding, only one major protein–DNA complex was observed in each case (Figure 6B, lanes 3 and 4). When both proteins were synthesized separately and mixed in the presence of DNA, again the two corresponding protein–DNA complexes were formed (lane 5). However, when mRNAs corresponding to the two proteins were cotranslated and the products were assayed for DNA binding, a complex of intermediate mobility was observed (lane 6). A similar result was also obtained when both proteins were translated separately and incubated together prior to the addition of the DNA (data not shown). We interpreted the formation of this intermediate complex as corresponding to a heterodimer formed between the full-length and the truncated version of GBF1. Similar results were obtained with GBF2 and GBF3 (data not shown).

The amino acid similarity between GBF1, GBF2 and GBF3 is not restricted to the basic region but extends into the leucine zipper domain. In addition to the conservation of each leucine residue (L1–L6), with one exception, every fourth position (a–f) interdigitating the heptad repeats is also conserved (Figure 3A). These amino acids are thought to be involved in the hydrophobic interactions between two leucine zipper domains and are required for dimerization (Hu *et al.*, 1990). This observation prompted us to investigate

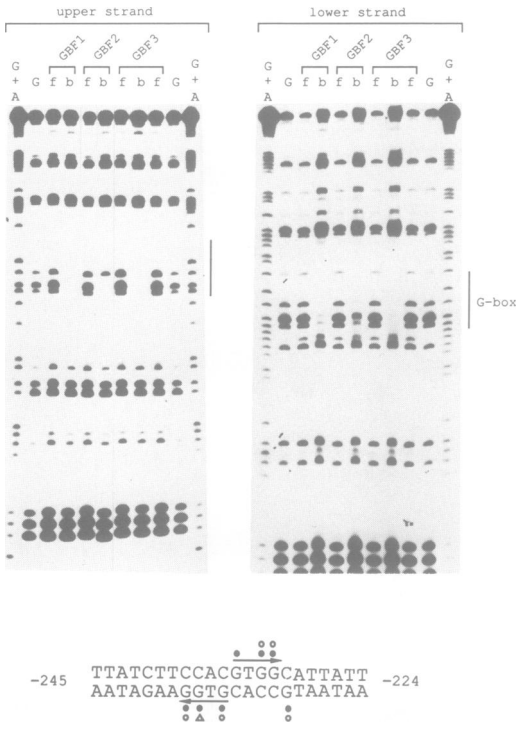


Fig. 5. Methylation interference studies. The partially methylated *Arabidopsis rbcS-1A* promoter fragment (−251 to −211) was incubated with either *in vitro* generated GBF1 (amino acids 1−315), GBF2 (amino acids 1−360) or GBF3 (amino acids 95−324), as indicated. After separation of the protein−DNA complexes, the DNA was cleaved with piperidine and the cleavage products were analyzed on an 8% sequencing gel. A+G and G represent Maxam−Gilbert (Maxam and Gilbert, 1980) sequencing reactions of the same DNA fragment. Bars indicate the location of the G-box. The DNA sequence of this DNA region is given below. Arrows mark the palindromic G-box. Closed circles indicate the G residues which, when methylated, interfere with GBF1 and GBF3 binding; open circles those which, when methylated, interfere with GBF2 binding. Methylation of the G residue marked by the triangle only partially inhibits GBF2 binding. The numbers refer to the location of this DNA element within the *Arabidopsis rbcS-1A* promoter.

whether the three proteins would form heterodimers and whether these heterodimers would also interact with the G-box element. Therefore, different combinations of full-length and truncated proteins (Figure 6A) were assayed for binding to the radiolabeled G-1A oligonucleotide. As illustrated in Figure 6C, the full-length GBF1 (GBF1L, lane 3) and truncated GBF3 (GBF3S, lane 4) proteins gave rise to distinct protein−DNA complexes which migrated with different mobilities. However, when the two proteins were mixed and incubated prior to the addition of the DNA, a complex of intermediate mobility was observed (lane 5). Similar results were obtained with GBF2L/GBF1S (Figure 6D, lane 5) and GBF2L/GBF3S (Figure 6E, lane 5). These intermediate protein−DNA complexes are interpreted as corresponding to heterodimers formed between the different GBF proteins. These results demonstrate that GBF1, GBF2 and GBF3 are able to form heterodimers *in vitro*. Moreover, all heterodimers interact with the *Arabidopsis rbcS-1A* G-box motif. Currently we do not know why none or very little of the GBF1L/GBF1L (Figure 6, panel B, lane 6 and panel C, lane 5) or the GBF3S/GBF3S (panel E, lane 5) homodimers are formed. More detailed analyses are required to determine the kinetics of homo- versus heterodimer formation.

The expression of GBF3 is light-regulated in contrast to that of GBF1 and GBF2

Southern blot analyses employing *Arabidopsis* genomic DNA revealed a simple hybridization pattern for each of the three cDNAs (data not shown). Furthermore, no cross-hybridization was obtained between the three cDNAs under the highly stringent hybridization conditions employed, indicating that these probes could be used as gene specific hybridization probes for the corresponding transcripts.

To determine the expression characteristics of *GBF1*, *GBF2* and *GBF3*, Northern analyses were carried out on RNA isolated from *Arabidopsis* leaves and roots. The leaf RNA was prepared from 5 day-old seedlings which were grown either under standard daylight conditions or in constant darkness. As shown in Figure 7, all three cDNAs hybridized to a mRNA species of almost identical size (1.5−1.6 kb). These results, in combination with our sequence data, demonstrated that the *GBF1* and *GBF2* cDNAs are nearly full-length and that 100−200 bp in the 5′ end of the GBF3 cDNA are missing. Furthermore, *GBF1* and *GBF2* mRNAs were present at approximately equal levels under the conditions examined; *GBF1* mRNA being present at slightly lower levels in light grown leaf tissue (Figure 7). These results indicated that *GBF1* and *GBF2* are expressed in both photosynthetically active and non-active tissue. In contrast, the hybridization pattern of the *GBF3* cDNA revealed that this gene is far more strongly expressed in leaves of dark grown plants, indicating that the expression is down-regulated by light. A longer exposure of the autoradiogram revealed slight expression in light grown tissue (data not shown). The highest *GBF3* mRNA levels were detected in roots, suggesting that the expression of this gene is also regulated in a tissue specific manner.

Discussion

Arabidopsis thaliana genome encodes several G-box binding proteins

The G-box motif (CCACGTGG) of the *Arabidopsis rbcS-1A* promoter is bound by the nuclear factor GBF. However, G-boxes and G-box-like elements seem to be required for a large array of responses in plants (Block *et al.*, 1990; Donald and Cashmore, 1990; Marcotte *et al.*, 1989; Mundy *et al.*, 1990; Schulze-Lefert *et al.*, 1989). These observations raise the possibility that several GBF-like proteins exist in a single plant species and that these proteins may be involved in different signal transduction pathways. Here we describe the isolation of three *Arabidopsis* cDNA clones encoding GBF proteins. The deduced amino acid sequence of the three cDNAs revealed that the proteins, designated GBF1, GBF2 and GBF3, are characterized by a C-terminal DNA binding domain encompassing a basic region followed by a leucine zipper motif. All three proteins interact specifically with the G-box of the photoregulated *Arabidopsis rbcS-1A* promoter as well as the root specific *adh* promoter. The N-terminal domains of all three proteins are highly enriched in proline residues. Proline-rich regions found in other mammalian transcription factors have been shown to potentiate transcription (Mermod *et al.*, 1989; Williams and Tjian, 1991) and we have demonstrated that the proline-rich region

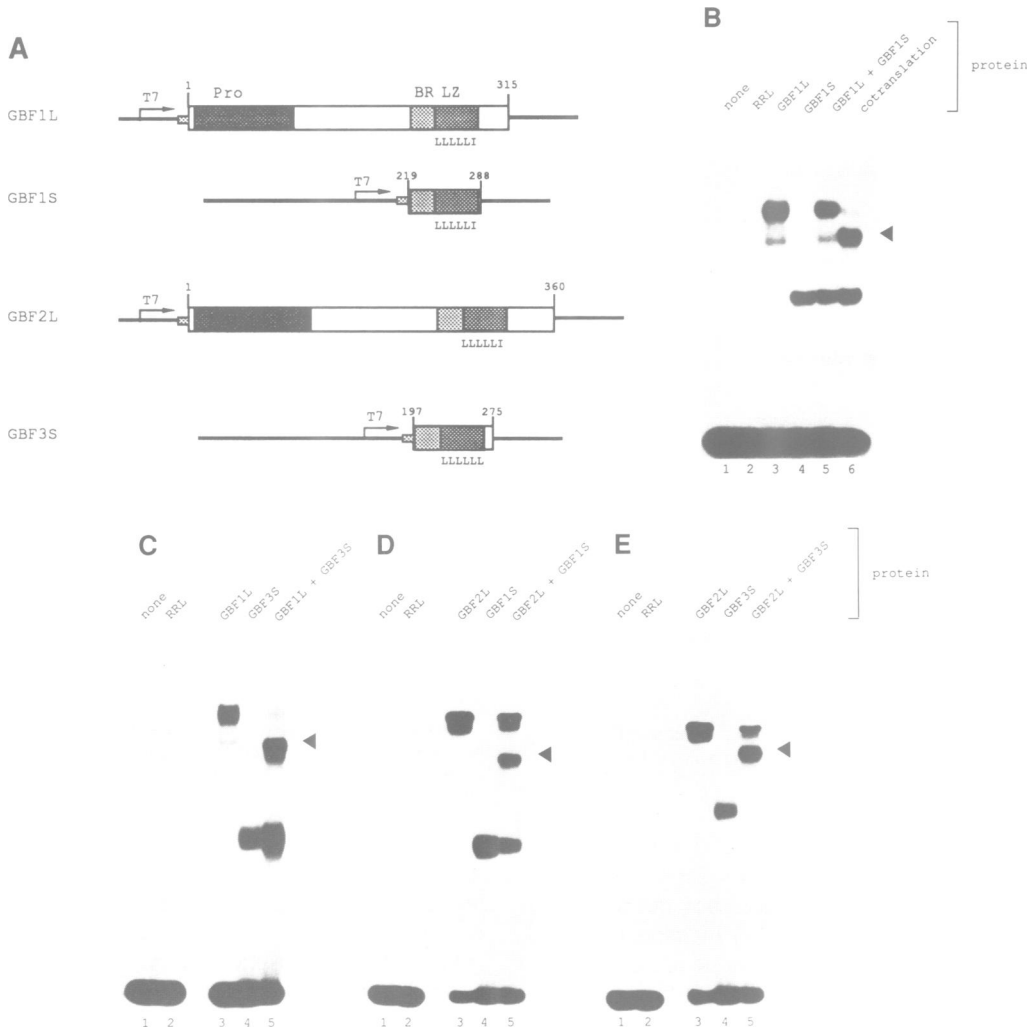


Fig. 6. GBF1, GBF2 and GBF3 heterodimerize promiscuously. (**A**) Schematic presentation of the templates used for generating full-length and truncated GBF1, GBF2 and GBF3 proteins. The proline-rich regions (Pro), basic regions (BR) and leucine zipper domain (LZ) are highlighted. Leucine (L) and isoleucine residues (I) within the LZ are indicated. Start and endpoint of each *in vitro* translation product are given above each construct. The location of the T7 promoter is indicated (T7). (**B**) Homodimer formation of GBF1. The radiolabeled G-1A oligonucleotides were incubated with either no protein (lane 1), rabbit reticulocyte lysate (RRL) without RNA (lane 2), the *in vitro* generated full-length protein GBF1L (lane 3) or the truncated version GBF1S (lane 4). For the binding reactions shown in the last two lanes, the two proteins GBF1L and GBF1S were either translated separately and mixed in the presence of DNA (lane 5) or were cotranslated (lane 6). (**C**–**E**) Heterodimer formation between GBF1, GBF2 and GBF3. Within each panel, lane 1 contains the free G-1A oligonucleotide and lane 2 contains the binding reaction with rabbit reticulocyte lysate (RRL) only. Lanes 3 and 4 represent the individual protein–DNA complexes obtained with the full-length or truncated proteins as indicated above each lane. For lane 5 of each panel, the indicated proteins were mixed and incubated prior to the addition of the DNA. The heterodimeric protein–DNA complexes of intermediate mobility are marked by closed triangles. The protein–DNA complexes were resolved on 5% (panels B and C) or 8% (panels D and E) non-denaturing polyacrylamide gels.

of GBF1 is capable of activating transcription in plant protoplasts and mammalian cells (Schindler *et al.*, 1992).

Structural similarity between the Arabidopsis GBF proteins and other plant bZIP proteins

Several plant bZIP proteins have been isolated from a variety of plant species. Based on their DNA recognition sequences, these proteins appear to fall into two somewhat overlapping classes: the TGACGT/C and the G-box (CCACGTGG) binding proteins. Although the two DNA motifs share a 4 bp core sequence (ACGT), proteins belonging to each class are distinct; this distinction is based not only on their binding properties but also on their overall protein structures. Proteins of the first class, including the tobacco protein TGA1a (Katagiri *et al.*, 1989), the wheat protein HPB-1b (Tabata *et al.*, 1991) and the maize protein OCSBF-1 (Singh *et al.*, 1990), are characterized by an N-terminal DNA binding domain. In contrast, all proteins constituting the second class—the G-box binding proteins—contain a C-terminal DNA binding domain; this class includes the wheat proteins EmBP-1 (Guiltinan *et al.*, 1990) and HBP-1a (Tabata *et al.*, 1989, 1991), the parsley proteins CPRF-1 and CPRF-3 (Weisshaar *et al.*, 1991) and the tobacco protein TAF-1 (Oeda *et al.*, 1991). Furthermore, most of these G-box binding proteins are characterized by an N-terminal, proline-rich, putative activation domain. The TGACGT/C binding proteins, on the other hand, contain acidic or glutamine-rich regions. In addition to these two classes of plant bZIP proteins, there is an intermediate class, such as the recently identified parsley protein CPRF-2, that recognize both a TGACGT/C and a G-box motif (Weisshaar *et al.*, 1991).

Fig. 7. Expression pattern of *GBF1, GBF2* and *GBF3* mRNAs. RNA blot analysis. Total RNA (40 μg per lane) was isolated from *Arabidopsis* leaves grown under standard day light (L) conditions or in complete darkness (D), and from roots (R). The RNA was hybridized to the three cDNAs encoding GBF1, GBF2 and GBF3. The filters were stripped and rehybridized to an *Arabidopsis* β-tubulin cDNA fragment (Marks *et al.*, 1987). The position of the 18S (1.8 kbp) ribosomal RNA is indicated.

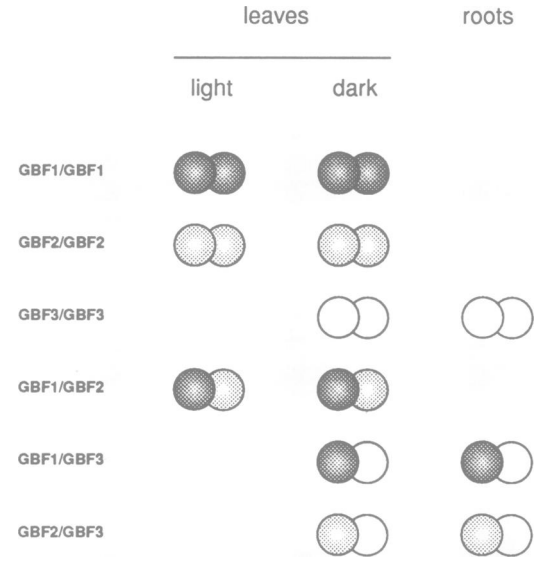


Fig. 8. Model explaining the differential distribution of homo- and heterodimeric GBF proteins in different *Arabidopsis* tissues and under different environmental conditions. Based on our Northern blot anlaysis we would predict that very little GBF3 is made in leaf tissue grown in the light, hence GBF1 and GBF2 homodimers and GBF1/GBF2 heterodimers are preferentially made. In dark-grown leaf tissue all three GBFs are expressed equally and therefore all six homo- and heterodimeric proteins could be present. In roots, GBF3 is strongly expressed and hence GBF3 homodimers and GBF1/GBF3 and GBF2/GBF3 heterodimers are preferentially made.

The properties of GBF1, GBF2 and GBF3 are similar in many ways to the other G-box binding proteins. The three *Arabidopsis* GBFs all interact with the G-box but not with the TGACGT/C motifs found in the as-1 element of the CaMV 35S promoter. Furthermore, the basic regions of the three proteins are more similar to other G-box binding proteins than they are to the TGACGT/C binding proteins. No similarity was observed outside the DNA binding domain between GBF1, GBF2 or GBF3 and the proteins belonging to the TGACGT/C binding class or to the parsley protein CPRF-2 of intermediate binding properties.

The DNA binding properties of GBF2 are somewhat distinct from those of GBF1 and GBF3

In reference to the question relating to the mechanism by which a single G-box element mediates distinct expression characteristics, we argued that it was important to determine whether there were single or multiple proteins interacting with this element. Having demonstrated the presence of multiple genes encoding GBF proteins, it then became of interest to determine whether this multiplicity reflected a redundancy in function or if the different proteins possessed distinct properties.

Observing differences in methylation interference patterns is one way to discriminate DNA binding properties of two proteins interacting with the same DNA motif. Our data clearly demonstrated that GBF2 interacted with the *rbcS-1A* G-box in a manner distinct from that of GBF1 and GBF3. Whereas methylation of all six G residues prevented binding of GBF1 and GBF3, binding of GBF2 was not inhibited by methylation of one of the central G residues (position −234, Figure 5). This data clearly demonstrated that there are distinct characteristics associated with the individual *A.thaliana* GBFs. Similar differences have been observed between the mammalian proteins ATF-1 and ATF-3 which both recognize the same DNA sequence and belong to the ATF/CREB transcription factor family (Hai *et al.*, 1989).

The methylation interference studies on GBF2 were also of interest in that they demonstrated a degree of asymmetry associated with the binding of this factor to the symmetrical G-box sequence. From this asymmetry we conclude that, after binding, this homodimeric protein is likely to possess directional information; i.e. one partner of the homodimer is distinct from the other This directional binding of the homodimer to a symmetrical palindromic sequence must reflect the asymmetry in the DNA sequences flanking the G-box. These flanking sequences may dictate asymmetry in the binding domain and may confer additional asymmetry elsewhere in the protein. This orientation dependent binding could be crucial for the interaction with other components of the transcriptional machinery. In this respect it is interesting to note that G-boxes or G-box-like elements are usually accompanied by other conserved DNA elements that are bound by different nuclear proteins. Within the *Arabidopsis rbcS-1A* promoter, for example, the G-box is flanked by two I-boxes that are essential for maximal expression levels and occupied by a nuclear factor (Donald and Cashmore, 1990; Schindler and Cashmore, 1990). Similarly, boxII (G-box) within the parsley *chs* promoter is not sufficient for light induced expression but requires the flanking boxI (Block *et al.*, 1990; Schulze-Lefert *et al.*, 1989; Weisshaar *et al.*, 1991). Interactions with such surrounding sequences may be facilitated by the asymmetry in binding that we have illuminated in our methylation interference studies on GBF2.

Relationship between the Arabidopsis GBF family and G-box binding proteins found in other plant species

A sequence comparison between GBF1, GBF2 and GBF3 revealed that within the N-terminal proline-rich region, GBF2 and GBF3 were more similar to each other (55 out of 92 amino acids) than to GBF1 (27 out of 68 and 20 out of 45 amino acids, respectively). Furthermore, when compared with G-box binding proteins isolated from other plant species, GBF2 and GBF3 were most similar to the parsley protein CPRF-1 (53 out of 114 and 42 out of 92 amino acids, respectively). However, GBF2 and GBF3 seem to be distinct from CPRF-1 based on their expression pattern. Weisshaar *et al.* (1991) demonstrated that, upon irradiation, *CPRF-1* mRNA accumulates in parsley suspension culture cells. In contrast, similar amounts of *GBF2* mRNA were detected in dark and light grown *Arabidopsis* leaves, indicating that the expression of *GBF2* is not light regulated. In contrast, the expression of *GBF3* was reduced upon irradiation. Whereas *GBF3* mRNA was strongly detected in dark grown leaves, very little mRNA was identified in light grown leaves. Similar observations were made with RNA isolated from adult *Arabidopsis* plants (data not shown). Furthermore, *GBF3* mRNA was most strongly expressed in roots. Based on these observations it appears that GBF2 and GBF3 are somewhat distinct from CPRF-1, although their N-terminal domains are very similar.

The N-terminal proline-rich domain of GBF1 is most similar to the wheat protein HBP-1a (Tabata *et al.*, 1989; 34 out of 64 amino acids) and the parsley protein CPRF-3 (Weisshaar *et al.*, 1991; 39 out of 64 amino acids). Here also the expression pattern of GBF1 is distinct from that identified for CPRF-3. Whereas *GBF1* mRNA is found in etiolated and light grown *Arabidopsis* leaves, the level of *CPRF-3* mRNA decreased upon irradiation (Weisshaar *et al.*, 1991). We do not know whether the similarity between GBF1 and HBP-1a extents to the expression characteristics as these have not yet been reported for HBP-1a.

Neither GBF1, GBF2 nor GBF3 share extensive sequence similarity with the N-terminal domain of the wheat G-box binding protein EmBP-1 (Guiltinan *et al.*, 1990; R.Quatrano, personal communication). The N-terminal domain of the tobacco protein TAF-1 (Oeda *et al.*, 1991) is somewhat similar to GBF3; however, the complete amino acid sequence of TAF-1 has not been determined yet. Interestingly, both *GBF3* and *TAF-1* mRNAs are most strongly expressed in roots (Oeda *et al.*, 1991). It will be interesting to see whether *TAF-1* mRNA also accumulates in etiolated tobacco leaves as observed for *GBF3* in etiolated *Arabidopsis* leaves.

Heterodimer formation increases the diversity of functional G-box binding proteins

In this study we have demonstrated that the number of functional G-box binding proteins in *Arabidopsis* can be increased by formation of heterodimers between the different GBF proteins. The heterodimers formed between GBF1, GBF2 and GBF3 interacted with the *rbcS-1A* G-box *in vitro*. Heterodimer formation has been characterized both *in vitro* and *in vivo* for other bZIP proteins (Williams *et al.*, 1991; Ziff, 1990 and references therein) and it seems likely that such heterodimers between GBF1, GBF2 and GBF3 will also form *in vivo*. These heterodimers, even when interacting with the same DNA motif, may constitute functional transcription factors with activation characteristics quite distinct from those of the original homodimers. Hence the activity of a G-box-containing target gene may be controlled by the subtle changes in the populations of GBF1, GBF2 and GBF3 within a given cell. Based on our level of analyses it appears that different amounts of individual homo- and heterodimers are present in roots or light and dark grown leaves (Figure 8). All three genes are expressed to approximately the same extent in dark grown leaves and hence, if we assume co-expression in the same cell and equal kinetics for the formation of all homo- and heterodimers, then similar amounts of six different proteins would be formed. In contrast, almost no *GBF3* mRNA is present in light grown leaves, therefore GBF1 and GBF2 homodimers or GBF1/GBF2 heterodimers will be overrepresented under these environmental conditions (Figure 8). On the other hand, *GBF3* mRNA is strongly expressed in roots and hence GBF3 homodimers or GBF3/GBF1 and GBF3/GBF2 heterodimers will be preferentially made in this tissue. These observations will help to explain how individual G-box-containing promoters are activated in different tissues and under different environmental conditions.

Materials and methods

Construction of an Arabidopsis cDNA expression library in λZAP

Using 5 μg polyadenylated mRNA from 3 day-old *Arabidopsis* seedling hypocotyls as template and oligo(dT) as primer, first strand cDNA synthesis was catalyzed by Moloney murine leukemia virus reverse transcriptase (Pharmacia). Second strand cDNA was made using the procedure of Gubler and Hoffman (1983) except that DNA ligase was omitted. After the second strand reaction, the ends of the cDNA were made blunt with Klenow fragment and *Eco*RI−*Not*I adaptors (Pharmacia) were ligated to each end. The cDNA was purified from unligated adaptors by spun-column chromatography using Sephacryl S-300 and size-fractionated on a 1% low melting point mini-gel. Size-selected cDNAs (0.5−1 kb, 1−2 kb, 2−3 kb, 3−6 kb) were removed from the gel using agarase (New England Biolabs), phenol−chloroform extracted and ethanol precipitated. A portion of each cDNA size fraction (0.1 μg) was co-precipitated with 1 μg of λZAP II (Stratagene) *Eco*RI-digested, dephosphorylated arms and then ligated in a volume of 4 μl overnight. Each ligation mix was packed *in vitro* using Gigapack II Gold packaging extract (Stratagene). Phage libraries ($0.25-3.5 \times 10^7$ recombinants) were obtained, a portion of which was screened with oligonucleotide probes.

Screening of the cDNA expression library

The cDNA expression library was screened for proteins specifically interacting with the G-3A oligonucleotide as described by Vinson *et al.* (1988) with the following modifications. The probe consisted of eight copies of the G-3A oligonucleotide, polymerized head-to-tail and was end-labeled with [α-^{32}P]dATP and Klenow. The nitrocellulose filters were treated for 10 min with buffer X (20 mM HEPES−KOH, pH 8, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT), supplemented with 6 M guanidinium hydrochloride. Renaturation was performed with buffer X. The filters were briefly rinsed in buffer Y (10 mM Tris−HCl, pH 7.5, 40 mM NaCl, 1 mM EDTA) and incubated for 1 h in the same buffer supplemented with 5% non-fat dry milk. The binding reaction was performed for 1 h in buffer Y containing the radiolabeled oligonucleotide (0.5×10^6 c.p.m./ml). The filters were washed for 20 min in buffer Y.

The cDNAs encoding GBF2 and GBF3 were isolated by screening the cDNA library with a DNA fragment encompassing the basic region of GBF1 (base pairs 848−985). The DNA probe was generated and radiolabeled by the polymerase chain reaction (PCR) using pGBF1 as template and two gene-specific primers extending from base pair 848 to 896 and from 956 to 985. In the radiolabeling reaction, the dCTP was substituted by [α-^{32}P]dCTP. The filters were prehybridized in 30% formamide, 5 × Denhardt's, 5 × SSPE, 10% dextran sulfate, 10 μg/ml salmon sperm DNA for 14 h at 42°C. The hybridization was performed for 24 h under the same conditions after adding the radiolabeled probe (5×10^6 c.p.m./ml). The

filters were washed twice for 30 min in 30% formamide, 0.5% SDS, 5 × SSPE at 42°C and twice for 30 min at 60°C.

DNA sequence analysis

Sequence analysis was performed on double-stranded DNA using the dideoxy chain termination reaction (Sanger *et al.*, 1977) and suitable subclones of the cDNA inserts or employing specific internal primers.

Oligonucleotides

All oligonucleotides were synthesized on an Applied BioSystems DNA synthesizer 380B. Oligonucleotides employed in DNA binding studies were synthesized with *Bgl*II and *Bam*HI overhangs and cloned into the *Bgl*II and *Bam*HI sites of pBgI (Donald *et al.*, 1990). The integrity of the DNA sequence was confirmed by sequencing.

Plasmids

All plasmids were constructed using standard techniques (Sambrook *et al.*, 1989). The *in vivo* excision, yielding recombinant plasmids pGBF1, pGBF1a, pGBF2 and pGBF3, was performed as recommended by the manufacturer (Stratagene). Templates used for the *in vitro* generation of the various GBFs were generated by PCR as follows: 5′ primers [carrying, in series, a T7 promoter sequence, the 5′-leader and ATG from the human β-globin gene (Norman *et al.*, 1988) and 20 nucleotides corresponding to the 5′ end of the three cDNAs] and 3′ primers (20 nucleotides corresponding to the 3′ end of the three cDNAs) were annealed to the plasmids pGBF1, pGBF2 or pGBF3. The exact DNA sequence of these primers is available upon request. A total of 25 PCR cycles was performed. The PCR products were precipitated and used directly for the *in vitro* transcription reaction.

Northern analysis

Total RNA was prepared from 5 day-old *A.thaliana* seedlings according to Ausubel *et al.* (1989). The seedlings were grown on Marashige and Skoog medium and kept either in a growth chamber (18 h light and 6 h dark) or in constant darkness. Root RNA was prepared from 3 week-old roots, grown in liquid culture (Gamborg's B5 medium). 40 μg RNA were separated on a 1.2% formaldehyde–agarose gel. Hybridization was performed at 42°C in 50% formamide, 5 × SSPE, 20 × Denhardt's, 0.1% SDS, 10 μg/ml salmon sperm DNA and radiolabeled probes (0.5 × 10^6 c.p.m./ml) corresponding to the three cDNAs were prepared by random hexamer labeling (Feinberg and Vogelstein, 1983). The filters were washed for 10 min in 5 × SSC and for 30 min in 0.5 × SSC at 65°C.

Preparation of crude plant extract

Nuclear extracts were prepared from 3 week-old *A.thaliana* and 8 week-old tobacco plants grown under greenhouse conditions. Proteins were extracted as described previously (Schindler and Cashmore, 1990).

Preparation of proteins from E.coli

Escherichia coli cells (DH5α) transformed with pGBF1 were grown to an OD_{600} of 0.5. Protein synthesis was induced by adding 2 mM IPTG and growth was continued for an additional hour. After centrifugation the cells were lysed in buffer X (see screening procedure) supplemented with 6 M guanidinium hydrochloride. The extracts were clarified by centrifugation. Renaturation of the proteins were performed by dialysis overnight against buffer X at 4°C.

In vitro transcription and translation

In vitro transcription was performed using 300 ng of the individual PCR products and T7 RNA polymerase as described by Melton *et al.* (1984). The RNA was translated in a 50 μl volume as recommended by the manufacturer's specifications (Promega).

Preparation of radiolabeled probes

The *rbcS-1A* promoter fragment −251 to −211 used for methylation interference experiments was prepared as described previously (Donald *et al.*, 1990; Schindler and Cashmore, 1990). Oligonucleotides or DNA fragments employed in DNA binding experiments were excised from the vector, radiolabeled by filling in the 5′ overhangs with [α-^{32}P]ATP and Klenow and gel purified.

Mobility shift assays and methylation interference experiments

The assays were performed as described previously (Schindler and Cashmore, 1990). The binding reactions employing *in vitro* translated protein contained 10^4 c.p.m. (~10 fmol) radiolabeled probe, 750 ng poly[d(I–C)], 0.1 pmol random single-stranded DNA and 1 μl *in vitro* translation product in a total volume of 15 μl. Unless indicated otherwise competitive binding assays were performed in the presence of 1 pmol of specific competitor DNA. Formation of heterodimers was investigated by incubating two different *in vitro* translation products (1 μl each) for 30 min at room temperature prior to the addition of the radiolabeled DNA binding site.

Acknowledgements

We thank Ralph Quartrano for providing us with unpublished sequence data. We are grateful to Margret Ahmad for synthesizing the oligonucleotide and reading the manuscript and to John Chan for useful discussions and helpful comments concerning the manuscript. This work was supported by NIH grant GM38409 to A.R.C. and by NIH grants GM38894 and HG00322 to J.R.E.

References

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1989) *Current Protocols in Molecular Biology*. Greene Publishing Associated/Wiley-Interscience.

Block,A., Dangl,J.L., Hahlbrock,K. and Schulze,L.P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5387–5391.

Cao,Z., Umek,R.M. and McKnight,S.L. (1991) *Genes Dev.*, **5**, 1538–1552.

Castresana,C., Garcia-Luque,I., Alonso,E., Malik,V.S. and Cashmore,A.R. (1988) *EMBO J.*, **7**, 1929–1936.

DeLisle,A.J. and Ferl,R.J. (1990) *Plant Cell*, **2**, 547–557.

Donald,R.G.K. and Cashmore,A.R. (1990) *EMBO J.*, **9**, 1717–1726.

Donald,R.G.K., Schindler,U., Batschauer,A. and Cashmore,A.R. (1990) *EMBO J.*, **9**, 1727–1735.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.

Ferl,R.J. and Laughner,B.H. (1989) *Plant Mol. Biol.*, **12**, 357–366.

Giuliano,G., Pichersky,E., Malik,V.S., Timko,M.P., Scolnik,P.A. and Cashmore,A.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7089–7093.

Gonzalez,G.A., Yamamoto,K.K., Fischer,W.H., Karr,D., Menzel,P., Biggs III,W., Vale,W.W. and Montminy,M.R. (1989) *Nature*, **337**, 749–752.

Gubler,U. and Hoffman,B. (1983) *Gene*, **25**, 263–272.

Guiltinan,M.J., Marcotte,W.R. and Quatrano,R.S. (1990) *Science*, **250**, 267–271.

Hai,T. and Curran,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3720–3724.

Hai,T., Liu,F., Coukos,W.J. and Green,M.R. (1989) *Genes Dev.*, **3**, 2083–2090.

Hoeffler,J.P., Meyer,T.E., Yun,Y., Jameson,J.L. and Habener,J.F. (1988) *Science*, **242**, 1430–1433.

Hu,J.C., O'Shea,E.K., Kim,P.S. and Sauer,R.T. (1990) *Science*, **250**, 1400–1403.

Katagiri,F., Lam,E. and Chua,N.-H. (1989) *Nature*, **340**, 727–730.

Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.

Lohmer,S., Maddaloni,M., Motto,M., Di Fonzo,N., Hartings,H., Salamini,F. and Thompson,R.D. (1991) *EMBO J.*, **10**, 617–624.

Marks,M.D., West,J. and Weeks,D.P. (1987) *Plant Mol. Biol.*, **10**, 91–104.

Marcotte,W.R., Russell,S.H. and Quatrano,R.S. (1989) *Plant Cell*, **1** 969–976.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.

McKendree,W.L., Paul,A.-L., DeLisle,A.J. and Ferl,R.J. (1990) *Plant Cell*, **2**, 207–214.

McKnight,S.L. (1991) *Sci. Am.*, **264**, 54–64.

Melton,D.A., Krieg,P.A., Ribagliati,M.R., Maniatis,T.P., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.

Mermod,N., O'Neill,E.A., Kelley,R.J. and Tjian,R. (1989) *Cell*, **58**, 741–753.

Mundy,J., Yamaguchi,S.K. and Chua,N.H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1406–1410.

Norman,C., Runswick,M., Pollock,R. and Teisman,R. (1988) *Cell*, **55**, 989–1003.

Oeda,K., Salinas,J. and Chua,N.-H. (1991) *EMBO J.*, **10**, 1793–1802.

O'Shea,E.K., Rutkowski,R. and Kim,P.S. (1989) *Science*, **243**, 538–542.

Pu,W.T. and Struhl,K. (1991) *Mol. Cell. Biol.*, **11**, 4918–4926.

Rasmussen,R., Benvegnu,D., O'Shea,E., Kim,P.S. and Alber,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 561–564.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning, A Laboratory Manual*, second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schindler,U. and Cashmore,A.R. (1990) *EMBO J.*, **9**, 3415–3427.
Schindler,U., Ecker,J.R. and Cashmore,A.R. (1991) In Jenkins,G.I. and Schuch,W. (eds), *Molcular Biology of Plant Development*. Society for Experimental Biology, Cambridge, Vol. XLV, pp. 211–218.
Schindler,U., Terzaghi,W., Beckmann,H., Kadesch,T. and Cashmore,A.R. (1992) *EMBO J.*, **11**, 1275–1289.
Schmidt,R.J., Burr,F.A., Aukerman,M.J. and Burr,B. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 46–50.
Schulze-Lefert,P., Dangl,J.L., Becker-Andre,M., Hahlbrock,K. and Schulz,W. (1989) *EMBO J.*, **8**, 651–656.
Singh,K., Dennis,E.S., Ellis,J.G., Llewellyn,D.J., Tokuhisa,J.G., Wahleithner,J.A. and Peacock,W.J. (1990) *Plant Cell*, **2**, 891–903.
Staiger,D., Kaulen,H. and Schell,J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6930–6934.
Staiger,D., Becker,F., Schell,J., Koncz,C. and Palme,K. (1991) *Eur. J. Biochem.*, **199**, 519–527.
Tabata,T., Takase,H., Takayama,S., Mikami,K., Nakatsuka,A., Kawata,T., Nakayama,T. and Iwabuchi,M. (1989) *Science*, **245**, 965–967.
Tabata,T., Nakayama,T., Mikami,K. and Iwabuchi,M. (1991) *EMBO J.*, **10**, 1459–1467.
Talanian,R., McKnight,C.J. and Kim,P.S. (1990) *Science*, **249**, 769–771.
Vinson,C.R., LaMarco,K.L., Johnson,P.R., Landschulz,W.H. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 801–806.
Vinson,C.R., Sigler,P.B. and McKnight,S.L. (1989) *Science*, **246**, 911–916.
Weisshaar,B., Armstrong,G.A., Block,A., Costa e Silva,O. and Hahlbrock,K. (1991) *EMBO J.*, **10**, 1777–1786.
Williams,T. and Tjian,R. (1991) *Genes Dev.*, **5**, 670–682.
Williams,S.C., Cantwell,C.A. and Johnson,P.F. (1991) *Genes Dev.*, **5**, 1553–1567.
Ziff,E.B. (1990) *Trends Genet.*, **6**, 69–72.

Received on October 28, 1991; revised on January 10, 1992

Note added in proof

The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X63894 (*GBF1*), X63895 (*GBF2*) and X63896 (*GBF3*).