The EMBO Journal vol.11 no.4 pp.1275-1289, 1992

DNA binding site preferences and transcriptional activation properties of the *Arabidopsis* transcription factor GBF1

Ulrike Schindler, William Terzaghi, Holger Beckmann[1], Tom Kadesch[1] and Anthony R.Cashmore

Plant Science Institute and Department of Biology, [1]Howard Hughes Medical Institute and Department of Human Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA

Communicated by M.van Montagu

The G-box is a *cis*-acting element found within the promoters of many plant genes where it mediates expression in response to a variety of different stimuli. This palindromic DNA motif (CCACGTGG) is composed of two identical half sites, the base pairs of which we have numbered −4 to +4 (numbering from 5′ to 3′). Both half sites are involved in the binding of the bZIP protein GBF1, a member of the GBF family of *Arabidopsis thaliana*. Here we demonstrate using the random binding site selection method that GBF1 interacts with, in addition to the palindromic G-box, other DNA motifs that fall into seven distinct groups. All groups share the ACGT core sequence, common to most DNA motifs bound by plant bZIP proteins so far characterized. Our studies demonstrate that a high affinity GBF1 binding site is further defined by the following two parameters: first, all sites contain a G residue at position +3 (as in ACGTG) and secondly, only certain base pair combinations are allowed at positions −4, −3 and +4. Two of the identified groups (TGACGTGG and TGACGTGT) contain the base pairs TG at positions −4 and −3 and hence resemble the binding sites of another class of plant bZIP proteins (TGACGT/C binding proteins). However, GBF1 only interacts with the TGACGT sequence if the two 3′ distal nucleotides (positions +3 and +4) are occupied by GG or GT. These data define the differences between a G-box binding protein and TGACGT/C binding proteins.

The N-terminal domain of GBF1 is defined by a high proline content. Such regions were also identified in proteins related to GBF1. We demonstrate that this N-terminal proline-rich domain of GBF1, when fused to a heterologous DNA binding domain, stimulates transcription in both plant protoplasts and mammalian cells. These extensive DNA binding studies and the characterization of the GBF1 activation domain will facilitate both the identification of regulatory elements and the *in vivo* function of GBF1.

Key words: Arabidopsis/bZIP protein/proline-rich transcriptional activation domain/GBF1 binding sites

Introduction

Transcriptional activation is primarily mediated through transcription factors that interact with enhancer and promoter elements. Binding of transcription factors to such DNA elements constitutes a crucial step in transcriptional initiation. Structural and functional analyses of transcription factors revealed that many of these proteins have a modular protein structure, i.e. they are composed of domains including a DNA binding and a transcriptional activation domain (Frankel and Kim, 1991; Johnson and McKnight, 1989; Latchman, 1990; Mitchell and Tjian, 1989). DNA binding domains have, in many cases, been well characterized. One of the better analyzed classes of DNA binding domain is the basic/leucine zipper (bZIP) motif composed of a basic region residing N-terminal to a leucine zipper region (Landschulz *et al.*, 1988). Whereas the basic region is required for specific protein−DNA interactions, the leucine zipper mediates homo- and heterodimerization (Hu *et al.*, 1990; Rasmussen *et al.*, 1991). In contrast to DNA binding domains, the structural architecture of transcriptional activation domains is less well defined. However, protein regions rich in acidic amino acids, glutamines or prolines have been demonstrated to activate transcription (Hope and Struhl, 1986; Ptashne, 1988; Williams *et al.*, 1988; Courey *et al.*, 1989; Mitchell and Tjian, 1989; Gerster *et al.*, 1990; Müller *et al.*, 1990). A common feature of these types of activation domains is that they augment transcription from a variety of promoters in different cell types (Mitchell and Tjian, 1989). Furthermore, they often function in a heterologous system and activate transcription when fused to another DNA binding domain (Ma *et al.*, 1988; Mermod *et al.*, 1989; Carey *et al.*, 1990; Katagiri *et al.*, 1990; Williams and Tjian, 1991).

Transcription factors are often encoded in small multigene families. Individual members of these families interact with identical or closely related DNA motifs that are found in a variety of promoters regulated by different stimuli or in a tissue specific manner (for review see Blackwell and Weintraub, 1990; Jones *et al.*, 1988; Weintraub *et al.*, 1991; Ziff, 1990). Recently we reported the isolation of three *Arabidopsis* bZIP proteins designated GBF1, GBF2 and GBF3. The three proteins interact with the G-box (CCACGTGG) of the photoregulated *rbcS-1A* (small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase) promoter and the root-specific *adh* (alcohol dehydrogenase) promoter. However, these proteins also recognize a similar element (aCACGTGG) found in the tomato *rbcS-3A* promoter (Schindler *et al.*, 1992). For the sake of convenience, we somewhat arbitrarily refer to this and other DNA sequences which differ by one or two base pairs from the palindromic G-box motif, as G-box-like elements. Families of related bZIP proteins have also been identified in other plant species, such as parsley and wheat. The three parsley proteins have been shown to interact with box II (G-box) of the photoregulated chalcone synthase (*chs*) promoter (Weisshaar *et al.*, 1991). A DNA binding protein from wheat, EmBP-1, interacts with a G-box-like element (aCACGTGG) found in the ABA-responsive element of the wheat Em promoter and with a related DNA sequence (tgACGTGG) of the wheat

histone 3 promoter (Guiltinan *et al.*, 1990). Another wheat protein, HBP-1a, initially designated HBP-1, also binds strongly to the tgACGTGG sequence (hereafter we will refer to this DNA motif in the context of the wheat histone 3 promoter as the histone 3 hexamer). In contrast, HBP-1a binds weakly to the parsley *chs* G-box (Tabata *et al.*, 1989, 1991). Only one G-box binding protein has been identified in tobacco so far (Oeda *et al.*, 1991). All these G-box binding proteins are very similar within the basic region; however, for the *Arabidopsis* GBF family we showed that they exhibit similar but slightly distinct DNA binding properties (Schindler *et al.*, 1992). A detailed analysis of the individual DNA binding properties will be required in order to understand at the molecular level the basis for any functional differences that may eventually be assigned to these different GBF factors.

The palindromic G-box motif (CCACGTGG) is composed of two identical half sites. Individual nucleotide positions within each half site are referred to as −4 to −1 and +1 to +4 (5′ to 3′). Based on methylation interference studies, both half sites appear to be required for binding of the *Arabidopsis* protein GBF1 (Schindler *et al.*, 1992). The 4 bp core sequence (ACGT, −2 to +2) is also found in the TGACGT motifs (as-1, ocs and nos) which are bound by other plant bZIP proteins (hereafter collectively referred to as TGACGT/C binding proteins; Katagiri *et al.*, 1989; Singh *et al.*, 1990; Tabata *et al.*, 1991). These TGACGT/C binding proteins are somewhat divergent from the G-box binding proteins described above (Schindler *et al.*, 1992). However, the TGACGT motif clearly overlaps with the wheat histone 3 hexamer sequence, which is bound by the wheat proteins HBP-1a and EmBP-1 that also interact with a G-box (Guiltinan *et al.*, 1990; Tabata *et al.*, 1989, 1991). Hence, there appears to be some relationship between G-box- and TGACGT/C-binding proteins. However, GBF1, for example, does not interact with the as-1 element of the CaMV 35S promoter, indicating that additional sequence information is required for GBF1 binding.

Previously we pointed out that G-box binding and TGACGT/C binding proteins differ not only in their DNA binding domains but also in their overall protein structures and their putative transcriptional activation domains (Schindler *et al.*, 1992). Transcriptional activation employing plant bZIP proteins has so far only been demonstrated for the tobacco TGACGT/C binding protein, TGA-1a, which is capable of stimulating transcription *in vitro* (Katagiri *et al.*, 1990; Yamazaki *et al.*, 1990), the tobacco protein TAF-1 (Oeda *et al.*, 1991) and the maize protein Opaque-2 (Lohmer *et al.*, 1991). The functional significance of the proline-rich regions found in most G-box binding proteins has not previously been reported.

As a first step towards discriminating between individual members of the *Arabidopsis* GBF family and to establish the relationship between different G-box- and TGACGT/C-binding proteins, we determined the exact binding site requirements for GBF1. Furthermore, we demonstrate that the proline-rich region of GBF1 can augment transcription in plant protoplasts as well as in mammalian cells. This determination of high affinity binding sites and the characterization of the GBF1 activation domain will contribute to the identification of the *in vivo* function of GBF1.

Results

Acidic amino acids flanking the basic residues are required for DNA binding of GBF1

The putative DNA binding domain of GBF1 is characterized by a basic region adjacent to the leucine zipper which mediates homo- and heterodimerization between *Arabidopsis* G-box binding proteins (Schindler *et al.*, 1992). The leucine zipper of GBF1 is composed of five leucine residues and one isoleucine residue regularly spaced at intervals of seven amino acids (Figure 1A). To characterize the boundaries of the DNA binding domain of GBF1 more precisely, we analyzed the binding properties of several truncated proteins using the electrophoretic mobility shift assay. Assays employed *in vitro* generated proteins (Figure 1A) and the G-1A oligonucleotide bearing the palindromic G-box of the *Arabidopsis rbcS-1A* promoter. As shown in Figure 1B, the full-length protein GBF1(1−315) strongly interacted with this DNA element (lane 3). The slightly more quickly migrating complex may reflect proteolytic breakdown of GBF1; however, we have no definitive evidence concerning the precise nature of this product. To determine whether the isoleucine of the leucine zipper is required for DNA binding, we generated GBF1(199−288). This protein not only lacks the C-terminus and part of the N-terminus but also carries an amino acid substitution (isoleucine to serine) within the last repeat of the leucine zipper (position 285, Figure 1A). This protein was still able to bind the G-box motif, indicating that the isoleucine is not required for homodimerization and hence is not crucial for DNA binding (Figure 1B, lane 4). In contrast, further 3′ deletions of the leucine zipper [GBF1(199−253), Figure 1A] completely abolished protein binding (Figure 1B, lane 5).

The basic region of GBF1 is composed of two clusters of basic amino acids: amino acids 224−228 (basic region A, BR-A) and amino acids 236−242 (basic region B, BR-B, Figure 1A). To determine the minimal amino acid requirements with this motif, various N-terminal truncated proteins were analyzed. The *in vitro* generated protein GBF1 (219−288) (Figure 1A) was still able to interact with the G-box (Figure 1B, lane 6). However, deletion of two additional amino acids [aspartate and glutamate, GBF1(221−288), Figure 1A] completely abolished the DNA binding activity (Figure 1B, lane 7). To investigate whether this loss of DNA binding activity was due simply to the truncation or whether the N-terminal amino acids, aspartate and glutamate (positions 219 and 220), were required for DNA binding, we substituted these two amino acids with alanines (Figure 1A). The resulting protein GBF1 (219_{AA}−288) also did not interact with the G-box (Figure 1B, lane 8). Separation of *in vitro* ^{35}S-labeled proteins by SDS−PAGE confirmed that equal amounts of these various proteins were used (data not shown). From these results we conclude that firstly, the five amino acids N-terminal to the basic residues of BR-A are required for DNA binding and secondly, the aspartate and glutamate residues at positions 219 and 220 play a crucial role in this protein−DNA interaction.

GBF1 binds to the non-palindromic G-box-like element of the Nicotiana plumbaginifolia cab-E promoter

GBF1 interacts strongly with the G-box (CCACGTGG) found in the *Arabidopsis rbcS-1A* promoter. Previously we

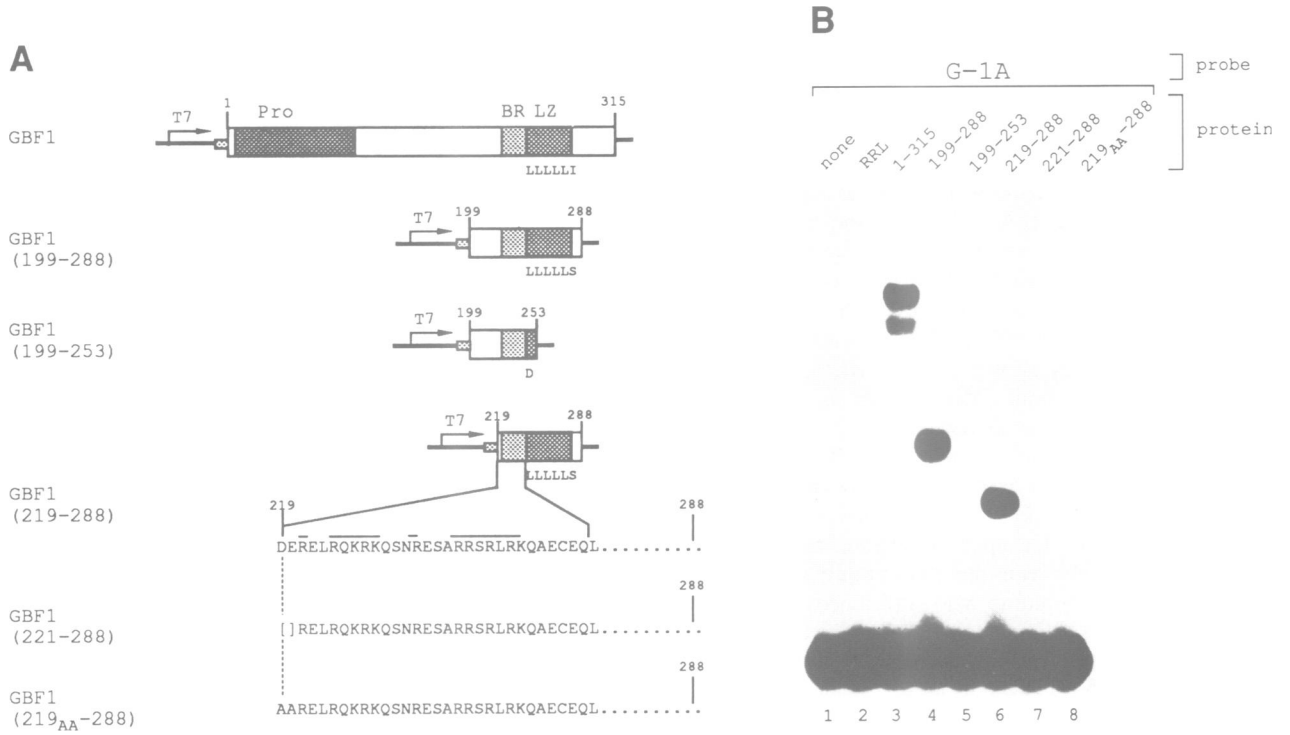


Fig. 1. Amino acids flanking the basic residues are required for DNA binding of GBF1. (**A**) Schematic presentation of constructs used for *in vitro* transcription and translation reactions. The location of the T7 promoter is indicated (T7). The proline-rich region (Pro), basic region (BR) and leucine zipper domain (LZ) of GBF1 are highlighted. The most N- and C-terminal amino acids of each *in vitro* translation product are indicated. The mutations introduced into the LZ are shown. GBF1(219−288), (221−288) and (219_{AA}−288) differ only within the N-terminal two amino acids. (**B**) Mobility shift assay. The labeled G-1A oligonucleotide derived from the *Arabidopsis rbcS-1A* promoter was incubated with no extract (lane 1), rabbit reticulocyte lysate (RRL) without RNA (lane 2) and *in vitro* translation products of the constructs outlined in panel A (lanes 3−8).

demonstrated that the G-box-like element (agACGTGG) of the *N.plumbaginifolia cab-E* promoter is required for maximal expression levels and is recognized by GBF present in crude nuclear extracts (Bringmann and Cashmore, unpublished data; Castresana *et al.*, 1988; Schindler and Cashmore, 1990). To determine if GBF1 exhibited these same binding characteristics, we performed mobility shift assays and showed that GBF1 bound to the *cab-E* promoter fragment (−301 to −186) carrying the G-box-like element (data not shown). We then carried out methylation interference studies employing the same promoter fragment and *in vitro* generated GBF1. Our results demonstrated that methylation of all G residues within this G-box-like motif (agACGTGG) prevented binding of GBF1 (Figure 2). The observed methylation interference pattern was identical to the one obtained with crude nuclear extracts (Schindler and Cashmore, 1990). These studies demonstrated that GBF1 binds to the *cab-E* G-box-like motif and that this binding is mediated by protein−DNA contacts similar to those observed with GBF.

GBF1 binds with high affinity to the tgACGTGG motif of the wheat histone 3 promoter

G-box motifs are found in several plant promoters. To investigate the influence of base pairs flanking this palindromic motif we carried out competitive DNA binding studies. Assays employed *in vitro* generated GBF1 and oligonucleotides derived from the *Arabidopsis rbcS-1A* (G-1A), *Arabidopsis* alcohol dehydrogenase (G-adh) and the parsley chalcone synthase (G-Pc) promoters (Figure 3A). As shown in Figure 3B, the complex formed between GBF1 and the G-1A oligonucleotide was strongly competed by the homologous oligonucleotide (lanes 4−6). The formation of this complex was abolished to the same extent by the G-adh oligonucleotide (lanes 7−9). Significant competition was also obtained with the oligonucleotide G-Pc (lanes 10−12). These data indicated that these naturally occurring DNA sequences carrying the palindromic G-box motif (CCACGTGG) are bound with similar affinities by GBF1.

The DNA binding domain of GBF1, especially basic regions A and B, is very similar (23 out of 25 amino acids) to the wheat bZIP protein HBP-1a (Figure 3D). HBP-1a was shown to interact strongly with the histone 3 hexamer (TGACGT) motif but only weakly with the G-box of the parsley chalone synthase promoter (Tabata *et al.*, 1991). This relatively weak binding of HBP-1a to the G-box sequence was somewhat surprising as the similarity between GBF1 and HBP-1a within the basic region suggested that both proteins may have closely related DNA binding specificities. To investigate directly whether GBF1 would exhibit DNA binding characteristics similar to those of HBP-1a, we compared the binding affinities of GBF1 with those of the G-1A oligonucleotide and the Hex oligonucleotide derived from the wheat histone 3 promoter (Figure 3A). As shown in Figure 3B, the Hex oligonucleotide competed to the same extent as the G-1A oligonucleotide for the complex formed between *in vitro* generated GBF1 and the G-1A oligonucleotide (lanes 13−15). A closer examination of the histone 3 hexamer site revealed that the TGACGT sequence is followed by two G residues and hence resembles

a G-box-like element (tgACGTGG). The strong binding of GBF1 to this non-palindromic histone 3 hexamer motif was rather unexpected; it is possible that the binding may have been facilitated by flanking DNA sequences. To investigate this hypothesis, the oligonucleotides Hexpp and Hex-1A were designed (Figure 3A). Hexpp carries the palindromic G-box in the context of the wheat histone 3 promoter and Hex-1A the sequence tgACGTGG in the context of the *rbcS-1A* promoter. As shown in Figure 3B, both oligonucleotides were bound by GBF1, Hex-1A slightly less efficiently than Hexpp (lanes 16–21). These data demonstrated that the G-box CCACGTGG and the sequence tgACGTGG are both substrates for GBF1 and that the binding affinity for both DNA motifs is only slightly influenced by flanking sequences.

As mentioned, HBP-1a—in contrast to GBF1—was shown to bind far more strongly to the histone 3 hexamer sequence than to the palindromic G-box of the parsley *chs* promoter (Tabata *et al.*, 1991). The apparent differences in the DNA binding affinities between the two proteins might be a reflection of the two amino acid differences within basic region A. Whereas GBF1 contains an arginine and glutamine residue at positions 225 and 230, respectively, HBP-1a carries a lysine and leucine at these corresponding positions (Figure 3D). To investigate whether the DNA binding affinity of GBF1 could be modulated simply by changing the arginine and glutamine residues to lysine and leucine respectively, the mutant protein GBF1-KL was generated (Figure 3D). This protein was assayed for DNA binding by the same procedure described for the wild-type GBF1 protein. As illustrated in Figure 3C, GBF1-KL exhibited DNA binding affinities identical to those observed for GBF1 (cf. Figure 3B). Although the DNA binding studies illustrated in Figure 3B and C were performed with a truncated GBF1 protein (219–288), similar results were obtained with the full-length wild-type and mutant proteins (data not shown). Our data demonstrated that GBF1, apparently unlike HBP-1a, binds with similar affinity to the non-palindromic tgACGTGG sequence and the palindromic G-box. Hence, it would appear that GBF1 and HBP-1a can be distinguished on the basis of their DNA binding affinities for G-box and some G-box-like elements. These apparent differences in DNA binding affinity are not due simply to the two amino acid differences with basic region A. However, comparative binding studies with both proteins are required to determine whether these apparent differences reside elsewhere in the protein sequences.

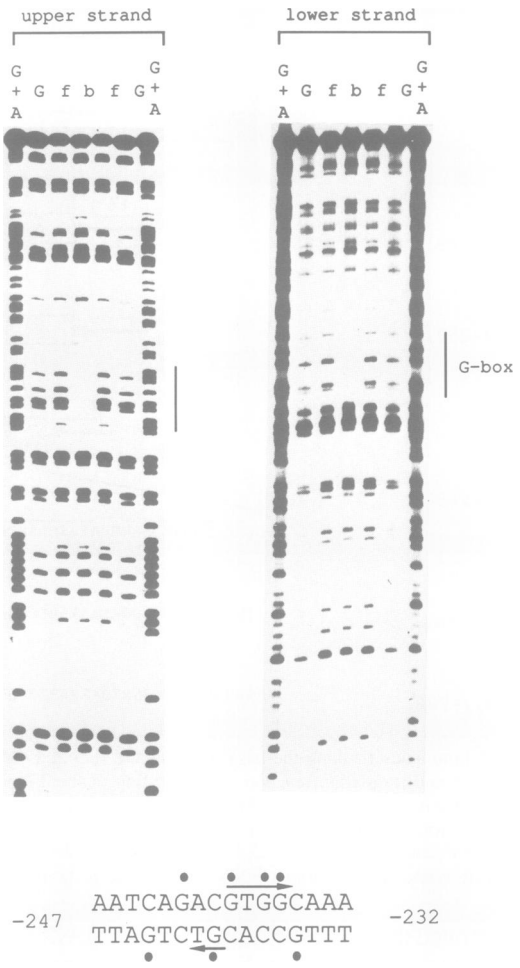
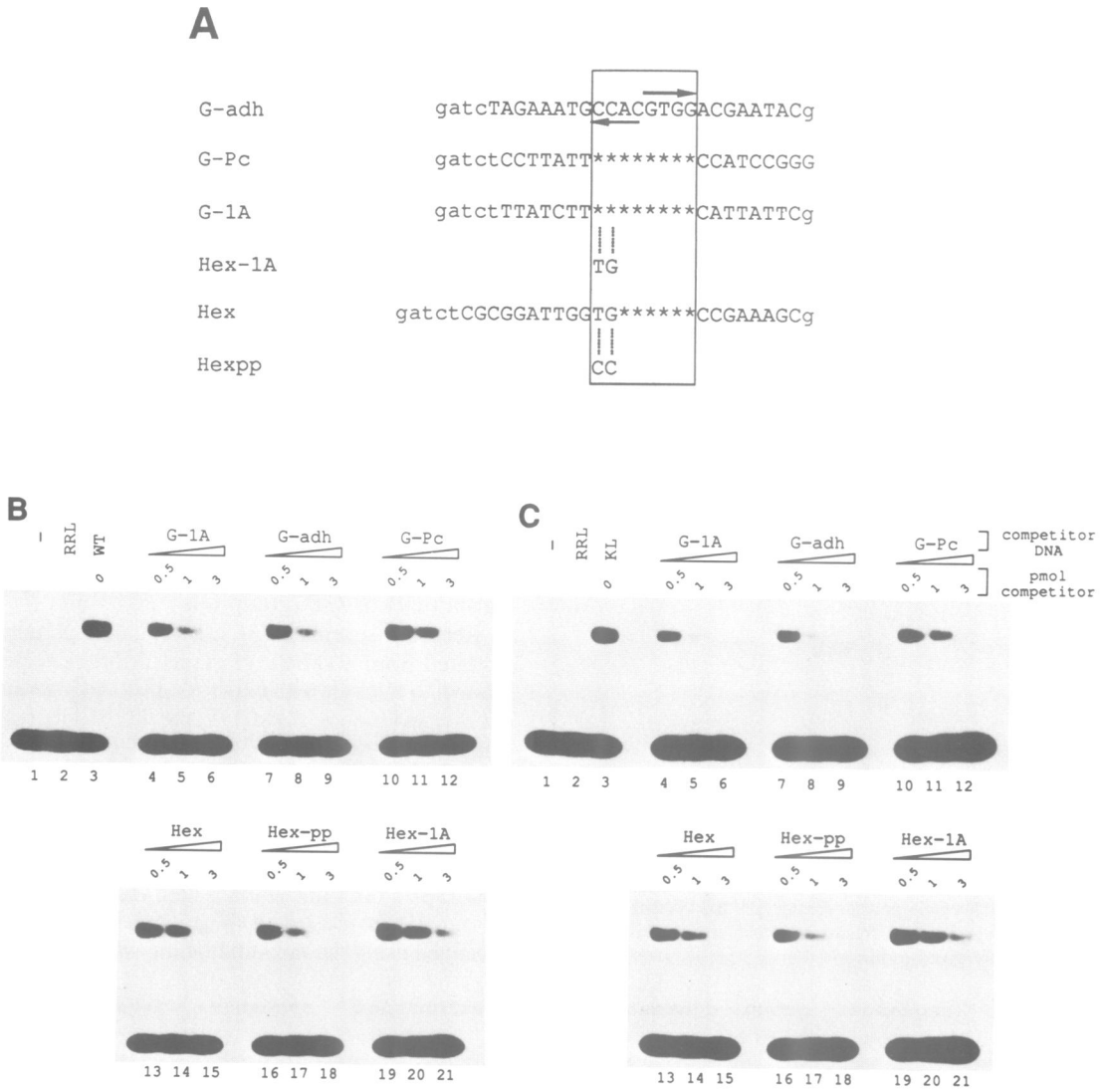


Fig. 2. GBF1 binds to the *cab-E* promoter element and the methylation interference pattern is identical to that observed with the nuclear factor GBF. Both strands of the *cab-E* promoter fragment (−301 to −186) were partially methylated and incubated with *in vitro* generated GBF1 protein. Free (f) and protein-complexed (b) DNA fragments were separated, eluted and, after piperidine cleavage, analyzed on a denaturing polyacrylamide gel. Markers (G and G+A) refer to Maxam–Gilbert sequencing reactions of this DNA fragment (Maxam and Gilbert, 1980). Bars indicate the location of the G-box. The DNA sequence of the protected region is given below. Arrows indicate the imperfect palindrome of the *cab-E* G-box-like element, filled circles designate the G residues which when methylated prevent protein binding.

Selection of high affinity binding sites for GBF1 from random oligonucleotides

The somewhat surprising observation that GBF1 bound with a similar affinity to both the palindromic G-box and the non-palindromic tgACGTGG (hex) motif of the wheat histone 3 promoter prompted us to carry out a more extensive study of the DNA binding site requirements of GBF1. We argued that such a detailed study of the DNA binding properties of GBF1—and eventually GBF2 and GBF3—would be necessary in order to define the functional properties of these closely related DNA binding proteins at the molecular level. Furthermore, our expectation was that such a study would shed light on the similarities and differences between G-box binding, hex binding and TGACGT/C binding bZIP proteins.

For these studies we used the random binding site selection method (Blackwell and Weintraub, 1990; Ekker *et al.*, 1991; Sun and Baltimore, 1991; Thiesen and Bach, 1990). This technique is based on the selection of specific protein binding sites from a pool of randomized DNA sequences and allows the identification of a large variety of potential target sequences. As a source of protein, GBF1 was overexpressed in *Escherichia coli* as a fusion protein containing six N-terminal histidine residues (Abate *et al.*, 1990). This protein, designated bGBF1, was purified to at least 95% homogeneity using affinity chromatography (Figure 4, lane 3). Binding sites were selected from a pool of oligonucleotides carrying 14 random base pairs flanked by 24 bp on either side (see legend to Figure 5). After five cycles of selection the bound oligonucleotide pool was cloned and the subclones were assayed individually for protein binding using the mobility shift assay and *in vitro* generated GBF1 protein (data not shown). DNA binding sites which were recognized by GBF1 with an affinity similar to the G-



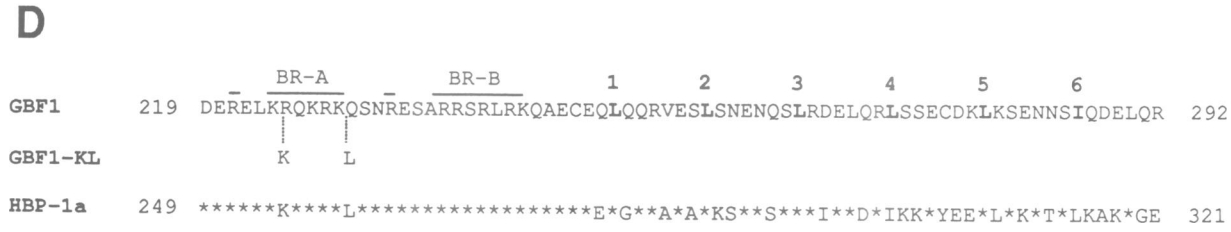


Fig. 3. The panlindromic G-box and the non-palindromic tgACGTGG motif are bound with equal affinities by GBF1. (**A**) DNA sequence of the oligonucleotides employed in competitive DNA binding studies. G-adh is derived from the *Arabidopsis adh* promoter (DeLisle and Ferl, 1990), G-PC from parsley *chs* promoter (Schulze-Lefert *et al.*, 1989). G-1A from the *Arabidopsis rbcS-1A* promoter (Donald and Cashmore, 1990) and Hex from the wheat histone 3 promoter (Tabata *et al.*, 1989). Base pairs identical to the CCACGTGG motif are indicated by asterisks. Hex-1A and Hexpp are mutant derivatives of the G-1A and Hex oligonucleotides respectively. (**B**) Competitive DNA binding studies employing wild-type GBF1. The radiolabeled G-1A oligonucleotides was incubated with no proteins (lane 1), rabbit reticulocyte lysate only (lane 2) or the *in vitro* generated wild-type GBF1(219−315) protein (lanes 3−21). Different amounts of various competitor DNAs, indicated above each lane, were included in the binding reactions (lanes 4−21). The DNA sequence of the competitor DNAs is given in panel A. (**C**) Competitive DNA binding studies employing mutant GBF1. The assays were performed as described for panel B except that the mutant protein GBF1(KL) was used. GBF1(KL) is identical to GBF1(219−315) except for the two amino acid exchanges indicated in panel D. (**D**) Amino acid comparison of the DNA binding domains of GBF1 and HBP-1a. The clusters of basic amino acids are overlined and designated basic regions A (BR-A) and B (BR-B). The leucines and isoleucines within the leucine zippers (bold letters) are numbered 1−6. Numbers on the left refer to the exact amino acid positions within both proteins. Identical amino acids are indicated by an asterisk.

box and a few of those which were bound with very low affinity were subsequently sequenced. A compilation of the selected sequences is shown in Figure 5. The sequences were divided into nine groups, designated I to IX, according to their consensus sequences. Individual positions within the consensus sequences were numbered −4 to −1 and +1 to +4 by analogy to the palindromic G-box sequence (group I). Sequences where non-random flanking base pairs appeared as part of the consensus sequence were omitted from each group, since statistically these sequences would be over-represented within the randomized pool. All DNA motifs within groups I−VIII share the ACGTG core sequence (positions −2 to +3), whereas the two clones of group IX carry base pair substitutions within this core sequence (Figure 5). DNA motifs which fall into groups I−VIII were bound with high affinity by either bGBF1 or *in vitro* translated GBF1 in mobility shift assays (data not shown) or competitive DNA binding studies (see below). DNA sequences within group IX were bound with substantially lower affinity (data not shown). From this random binding site selection assay we arrived at the following conclusions. Firstly, the core motif ACGTG (−2 to +3) is required for high affinity binding by GBF1. Secondly, only certain base pair combinations (CC, AC, TG, CT or TT) at positions −4 and −3 reconstitute a high affinity binding site. Thirdly, position +4 is occupied by either G or T. Fourthly, in agreement with our earlier results (Figure 3B), the histone 3 hexamer sequence in this particular context (tgACGTGG) is a high affinity GBF1 binding site. Fifthly, the sequence TGACGT (position −4 to +2), which represents the binding site of another class of plant bZIP proteins (TGACGT/C binding proteins), can be bound by GBF1 if the two following positions (+3 and +4) are occupied by either GG (group IV) or GT (group V).

Recently, several TGACGT/C binding proteins have been isolated from a variety of plants. Most of these have been shown to interact with either the histone 3 hexamer element or the two motifs TGACGT and TGACGC (as-1 element) identified in the CaMV 35S promoter and the ocs elements of the octopine and nopaline synthase promoters (Katagiri *et al.*, 1989; Singh *et al.*, 1990; Tabata *et al.*, 1991). The DNA binding domains of these proteins differ significantly from that identified for GBF1 and we have demonstrated that GBF1 does not interact with the as-1 motif (Schindler *et al.*, 1992). Based on these observations and the results obtained using the random binding site selection, it appeared

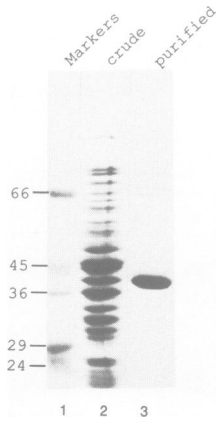


Fig. 4. Purification of bGBF1 from *E.coli*. The protein bGBF1 was overexpressed in *E.coli* and purified as described by Abate *et al.* (1990). Crude bacterial extracts (lane 2) and purified protein (lane 3) were analyzed on a 10% SDS−polyacrylamide gel and visualized by Coomassie blue staining. Lane 1 contains molecular mass standards. Numbers on the left are given in kDa.

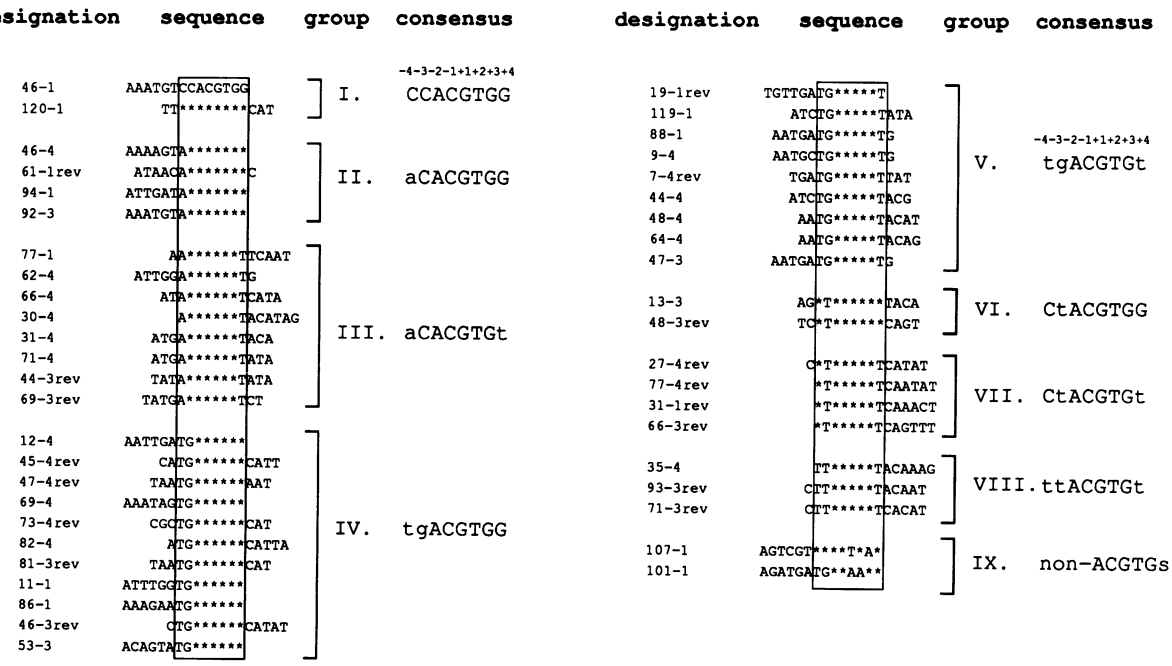


Fig. 5. Compilation of GBF1 DNA binding sites identified by the random binding site selection assay. DNA binding sites selected from a pool of oligonucleotides containing 14 random base pairs (5′-GTCTGTCTGAGGTGAGGGATCCTAN$_{14}$ACAAGCTTGTCTTCCGACGTCGCG-3′). Only the nucleotides corresponding to the random portion of the original oligonucleotides are shown. The identical binding sites were aligned and grouped according to their core sequences (boxed). The groups were designated I to IX. The individual positions within the eight base pair consensus sequences were designated −4 to +4. 'rev' refers to the reverse orientation of the individual sequences. Nucleotides identical to the G-box motif (CCACGTGG) are marked by asterisks.

that GBF1 bound only to those TGACGTXX motifs where the two base pairs XX were represented by GG (group IV) or GT (group V, Figure 5). To confirm this observation further, we performed protein binding studies employing two mutant derivatives (Hexm1 and Hexm2, Figure 6A) of the Hex oligonucleotide. The mutant sequence Hexm1 (TtA-CtTGG), which was originally used to verify the HBP-1a is a TGACGT/C binding protein, lacks both the TGACGT and the G-box-like element (Tabata *et al.*, 1989). In contrast, Hexm2 (TGACGTtt) contains two base pair substitutions within the G-box-like sequence, but it retains the integrity of the TGACGT sequence. The G-box-like element of the tomato *rbcS-3A* promoter (aCACGTGG, G-3A) and the as-1 oligonucleotides were used as positive and negative controls respectively (Figure 6A). As illustrated in Figure 6B, the wild-type Hex and G-3A oligonucleotides were the only sequences that were recognized by GBF1 (lanes 3 and 15). GBF1 did not interact with either of the Hex mutants (lanes 6 and 9) or with the as-1 element (lane 12). These data clearly demonstrated that binding of GBF1 is sensitive to the nature of the two 3′-terminal base pairs within the TGACGTXX motif.

Some G-box-like promoter elements are bound with low affinity by GBF1

Our random binding site selection assay revealed that GBF1 binds with high affinity to G-box-like sequences that carry certain base pair combinations in positions −4, −3 and +4 (Figure 5). G-box-like elements containing other base pair substitutions in these particular positions were not identified in this assay, but such sequences are found in a variety of plant promoters and several of these have been shown to be recognized by nuclear proteins. For example, we demonstrated that the G-box-like element of the *N.plumbaginifolia cab-E* promoter (agCAGTGG) was recognized by GBF1 (Figure 2). Other DNA elements that were not identified in the random binding site selection assay included motif I (gtACGTGG), a G-box-like element in the ABA inducible rice *rab16A* promoter that is bound by the tobacco protein TAF-1 (Oeda *et al.*, 1991). Also, the two G-box-like sequences (CCACGTat and tCACGTGc) of the *Antirrhinum majus chs* promoter were not isolated in this assay. The latter of the two elements was shown to be bound by the nuclear protein CG-1, which also interacts with the sequence CCACGTGa of the adenovirus major late promoter, originally identified as the binding site of the mammalian transcription factor USF (Sawadogo and Roeder, 1985; Staiger *et al.*, 1989). Similarly, the cold-induced *Arabidopsis Cor180* promoter also contains the promoter element atACGTGt (M.Thomashow, personal communication), which was not isolated. There are four reasons which could account for the absence of these DNA motifs from the collection of sequences obtained in the random binding site selection assay. First, GBF1 does not interact with these sequences. Secondly, these elements represent low affinity binding sites for GBF1. Thirdly, binding of GBF1 to these DNA motifs is strongly influenced by flanking base pairs and therefore it is statistically more unlikely that these sequences will be identified in random binding site selection assays. Fourthly, the compilation of high affinity GBF1 binding sites identified by the random binding site selection assay may be incomplete. To address these questions directly, we compared the binding affinity of GBF1 for these naturally occurring G-box-like elements with those identified in the random binding site selection assay using competitive DNA binding studies. The three oligonucleotides that were randomly chosen from our collection corresponded to groups III, IV and VII and are illustrated in Figure 5. Figure 7B shows that these sequences, 66-4 (lanes 4−7), 47-4 (lanes 8−11) and 27-4 (lanes 12−15) were able to compete to the same extent as the palindromic G-box (lanes 16−19) for the complex formed between GBF1 and G-1A. These

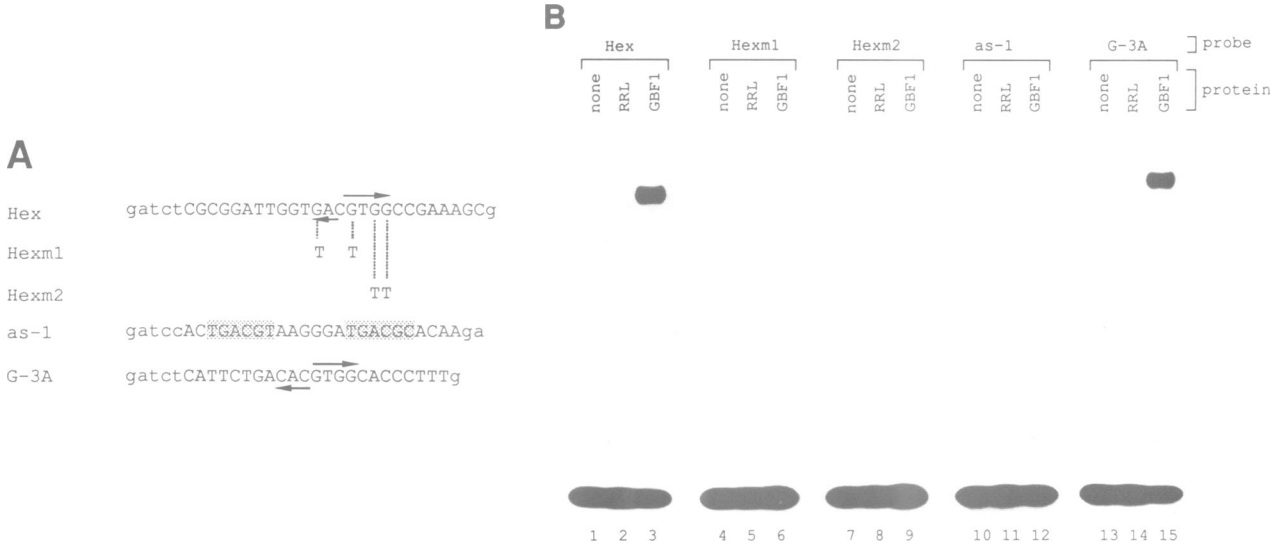


Fig. 6. GBF1 is sensitive to the nature of the two base pairs (XX) found 3′ distal within the TGACGTXX motifs. (**A**) DNA sequence of oligonucleotides employed in DNA binding studies. The Hex oligonucleotide is derived from the wheat histone 3 promoter. Hexm1 and Hexm2 are mutant derivatives carrying the indicated base pair substitutions. G-3A carries the G-box-like element of the tomato *rbcS-3A* promoter. as-1 contains the TGA-1a binding sites of the CaMV 35S promoter. Arrows indicate the location of the G-box-like elements. The TGACGT/C motifs are highlighted. (**B**) Electrophoretic mobility shift assay using radiolabeled oligonucleotides indicated above each panel. The reactions contained either no extract (lanes 1, 4, 7, 10 and 13), rabbit reticulocyte lysate (RRL) without RNA (lanes 2, 5, 8, 11 and 14) or *in vitro* generated GBF1 (lanes 3, 6, 9, 12 and 15). The DNA sequence of the oligonucleotides is given in panel A.

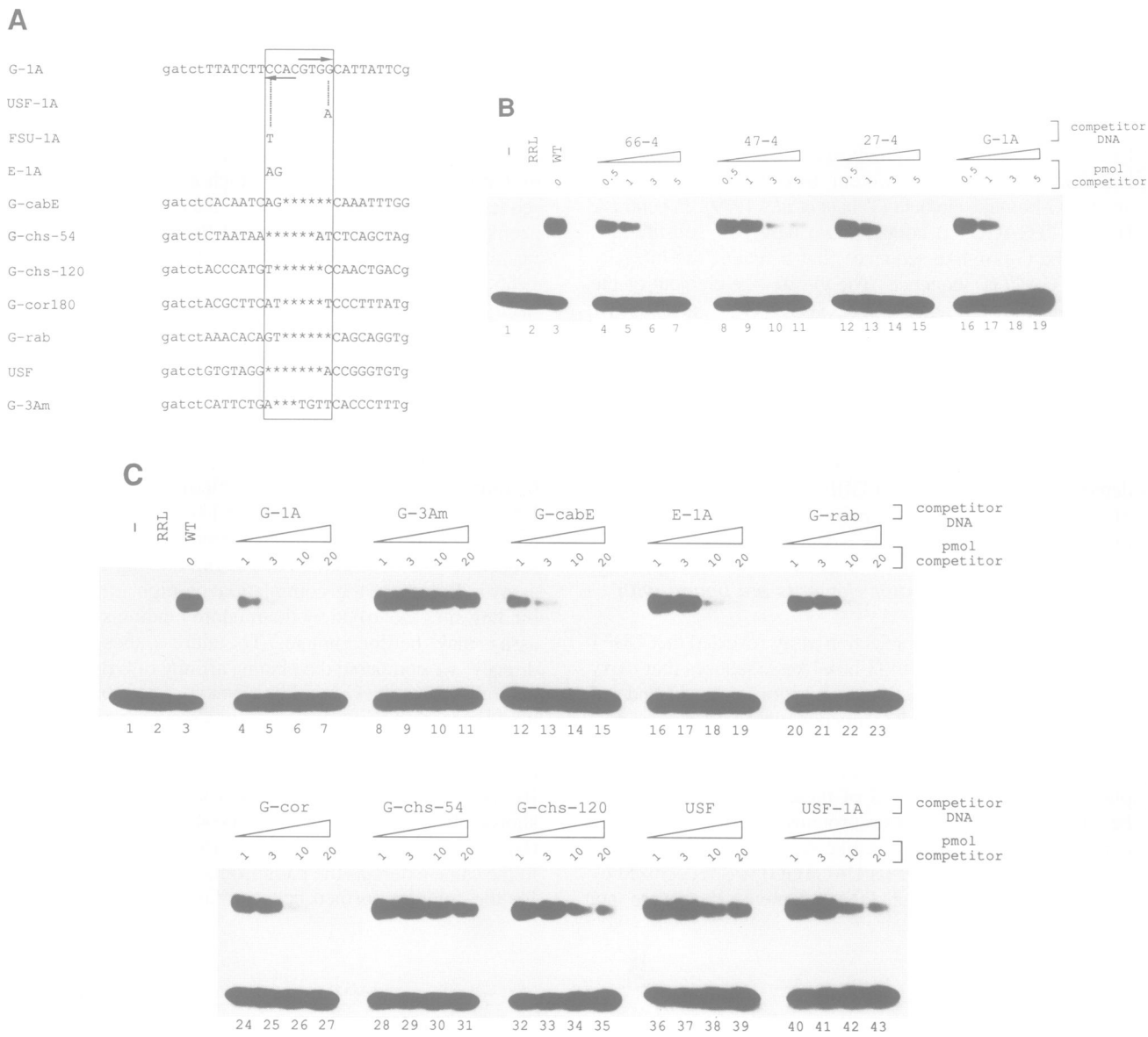
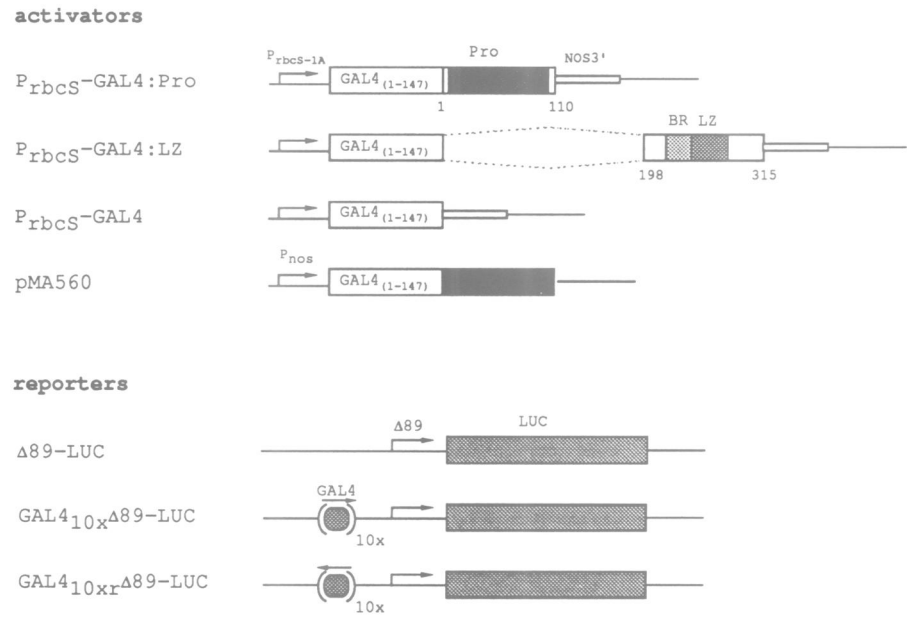


Fig. 7. Some naturally occurring G-box-like elements are bound with low affinity by GBF1. (**A**) DNA sequence of oligonucleotides derived from various promoters. G-1A carries the *Arabidopsis rbcS-1A* G-box (Donald and Cashmore, 1990). The following three G-1A derivatives represent single or double base pair substitutions introducing the USF element (USF-1A), the reverse orientation of the same element (FSU-1A) or the *N.plumbaginifolia cab-E* G-box-like element (E-1A) into the context of the *rbcS-1A* promoter. G-cabE represents the natural context of the G-box-like element found in the *N.plumbaginifolia cab-E* promoter (Schindler and Cashmore, 1990). The two elements G-chs-54 and G-chs-120 are found in the *A.majus chs* promoter (Staiger *et al.*, 1989); G-cor180 in the cold-induced *Arabidopsis cor180* promoter (M.Thomashow, personal communication); G-rab in the rice *rab16A* promoter (Oeda *et al.*, 1991) and USF in the adenovirus major late promoter (Gregor *et al.*, 1990). G-3Am is a mutant derivative of the tomato *rbcS-3A* G-box-like element (see Figure 6A). Arrows mark the palindromic G-box. The G-box and G-box-like motifs are boxed; nucleotides identical to the G-box are indicated by asterisks. (**B**) Competitive DNA binding assay employing 'randomly' selected oligonucleotides. The radiolabeled G-1A oligonucleotide was incubated with either no protein (lane 1), rabbit reticulocyte lysate (RRL, lane 2) or *in vitro* generated GBF1(219−315) (lanes 3−19). The binding reactions shown in lanes 4−15 contain different amounts of various competitor DNAs as indicated above the lanes. The DNA sequences of the competitor DNAs are illustrated in Figure 5. (**C**) Competitive DNA binding assay employing naturally occurring G-box-like elements and some mutant derivatives. Assays were performed as described for panel B, except that higher amounts of competitor DNAs, indicated above each lane, were used. The DNA sequences of the competitor DNAs are shown in panel A.

data demonstrated that we indeed selected for high affinity GBF1 binding sites in the random binding site selection assay. In contrast, the G-box-like element of the *N.plumbaginifolia cab-E* promoter (G-cabE, Figure 7A) was bound slightly less efficiently; about three times more competitor DNA was required than with the palindromic G-box (Figure 7C, lanes 4−7 and 12−15). Furthermore, binding of GBF1 to the *cab-E* sequence was shown to be strongly context dependent. The oligonucleotide E-1A carrying the *cab-E* G-box-like sequence (agACGTGG, Figure 7A) in the context of the *Arabidopsis rbcS-1A* promoter was bound with ~3−4 times lower affinity than the natural *cab*-E motif (Figure 7C, lanes 16−19). This result suggests that the *cab-E* box-like sequence is bound more efficiently in its natural context; this would result in a statistical reduction in the chance to identify such sequences from a randomized oligonucleotide pool.

It is of interest to contrast the above result, where the

A



B

reporter / activator	Δ89-LUC mean±SE	Δ89-LUC fold induction	Gal4$_{10x}$Δ89-LUC mean±SE	Gal4$_{10x}$Δ89-LUC fold induction	Gal4$_{10xr}$Δ89-LUC mean±SE	Gal4$_{10xr}$Δ89-LUC fold induction
None	10.61±2.34		2.51±0.42		1.76±0.45	
prbcS-GAL4:Pro	9.50±1.77	0.90	18.53±3.32	7.3	14.18±4.35	8.1
prbcS-GAL4:LZ	8.98±1.27	0.85	2.27±0.43	0.91	1.82±0.42	1.03
prbcS-GAL4	9.64±1.88	0.91	2.71±0.69	1.08	1.84±0.44	1.05
pMA560			29.19±2.17	11.6		

Fig. 8. The proline-rich region stimulates transcription in plant protoplasts. (**A**) Different portions of the GBF1 coding region were fused to the GAL4 DNA binding domain ($GAL4_{1-147}$). The functional domains of GBF1 are indicated (Pro, proline-rich region; BR, basic region; LZ, leucine zipper). Numbers below the boxes denote amino acids corresponding to GBF1. P_{rbcS1A} designates the *rbcS-1A* promoter from *Arabidopsis*; P_{NOS} and NOS3′ represent the promoter and polyadenylation signal, respectively, of the nopaline synthase gene. The luciferase reporter gene (*LUC*) was driven by the truncated CaMV 35S promoter (Δ89) with or without 10 copies of the GAL4 DNA binding site, in the correct [$(GAL4)_{10}$Δ89-LUC] and reverse [$(GAL4)_{10r}$Δ89-LUC] orientations. (**B**) Results (means ± standard errors) from four independent experiments where the indicated activator and reporter constructs were coelectroporated into SB-M protoplasts. Luciferase activities extracted from cells harvested two days after electroporation are expressed in arbitrary units relative to CAT activities.

surrounding sequences clearly have an effect on the binding properties of the *cab-E* G-box, with our earlier conclusion that surrounding sequences have relatively little effect on the binding properties of the palindromic *Arabidopsis rbcS-1A* G-box. The not too surprising conclusion would appear to be that surrounding DNA sequences can have a significant effect on the binding characteristics of more weakly binding sequences—such as the *cab-E* G-box—but may have little effect on the ability of the strongly binding *Arabidopsis rbcS-1A* G-box to bind to GBF1.

The DNA motifs of the *Arabidopsis cor180* and rice *rab16A* promoters (G-cor and G-rab, Figure 7A) were also bound with the same low affinity as E-1A (Figure 7C, lanes 20−27). Furthermore, the two DNA motifs derived from the *A.majus chs* promoter (G-Am-54 and G-Am-120, Figure 7A) failed almost completely to compete for GBF1 binding (Figure 7C, lanes 28−35). The low affinity of GBF1 for these naturally occurring sequences explains why we did not identify those motifs in our selection assay.

CG-1, a nuclear factor that recognizes the sequences tCACGTGc (G-Am-120, Figure 7A) of the *A.majus chs* promoter was also shown to interact with the binding site of the mammalian factor USF (CCACGTGa, Staiger *et al.*, 1989). We have demonstrated that the USF binding site competed very little for binding of GBF present in crude nuclear extracts (U.Schindler and A.R.Cashmore, unpublished data). The same results were obtained with GBF1 (Figure 7C, lanes 36−39), supporting the argument that GBF and GBF1 have similar DNA binding properties and that they are distinct from the binding characteristics of CG-1. The USF binding site differs by only one base pair substitution (position +4) from the G-box. Furthermore, the reverse orientation of this sequence (tCACGTGG) is similar not only to the G-box (group I) but also to the tgACGTGG

motifs represented in group IV. To determine whether the lack of GBF1 binding to the USF binding site was due to this one base pair substitution within the recognition sequence or to flanking DNA sequences, two mutant oligonucleotides, USF-1A (CCACGTGa) and FSU-1A (tCACGTGG, Figure 7A), were designed. Both carry the USF binding site in either orientation in the context of the *rbcS-1A* promoter. Figure 7C illustrates that the USF-1A (lanes 40–43) failed almost completely to compete for binding of GBF1 to the G-1A oligonucleotide. The same results were obtained with FSU-1A (data not shown). These data indicate that a single base pair substitution (position −4 or +4) can have a dramatic effect on the binding of GBF1. Furthermore, these results provide a striking example of the sensitivity of binding of GBF1 to base pair combinations at positions −4 and −3 of the binding site. Whereas motifs containing CC, AC, TG, CT or TT at positions −4 and −3 represent high affinity binding sites (Figure 5 and 7), motifs flanked by TC, as found in the reverse USF element, represent very low affinity binding sites.

The proline-rich region of GBF1 mediates transcriptional activation in plant protoplasts

An interesting feature of the N-terminal domain of GBF1 is its high proline content (Schindler *et al.*, 1992). A number of transcription factors identified in mammalian cells (Gronemeyer *et al.*, 1987; Kumar *et al.*, 1987; Santoro *et al.*, 1988; Williams *et al.*, 1988) and *Drosophila* (Laughon and Scott, 1984; Rosenberg *et al.*, 1986) contain proline-rich regions. Furthermore, the plant bZIP proteins CPRF-1, CPRF-3 and HBP-1a which interact with a G-box or G-box-like elements also contain an N-terminal domain rich in proline residues (Tabata *et al.*, 1989; Weisshaar *et al.*, 1991). For the two mammalian transcription factors CTF-1 and AP-2 it was demonstrated that these regions are required for transcriptional stimulation (Mermod *et al.*, 1989; Williams and Tjian, 1991).

To determine directly whether the proline-rich region of GBF1 is capable of activating gene expression, we linked it to a heterologous DNA binding domain. As an appropriate system we chose the DNA binding domain of the yeast transcription factor GAL4 ($GAL4_{1-147}$) since there is no endogenous GAL4-like activity in plant cells (Ma *et al.*, 1988). The different activator constructs are illustrated in Figure 8A. P_{rbcS}-GAL4:Pro is composed of the GAL4 DNA binding domain and the proline-rich N-terminal domain of GBF1 (amino acids 1–110). The construct P_{rbcS}-GAL4:LZ was used to investigate whether the C-terminal domain of GBF1 would also be able to augment transcription. This construct contains amino acids 198–315, encompassing the basic region and the leucine zipper of GBF1 attached to the DNA binding domain of GAL4. P_{rbcS}-GAL4 was used as a control plasmid. The expression of all three proteins was driven by the *rbcS-1A* promoter, which is strongly expressed in these protoplasts under the assay conditions used (W.Terzaghi and A.R.Cashmore, unpublished data). The activator constructs were then assayed for their ability to activate the expression of a heterologous template (Figure 8A); for this we chose the luciferase (*LUC*) gene fused to the truncated CaMV 35S promoter (Δ89). The three reporter constructs differed only in the presence [$(GAL4)_{10x}$−Δ89-LUC and $(GAL4)_{10xr}$−Δ89-LUC] or absence (Δ89-LUC) of ten GAL4 DNA binding sites.

Averages of four independent experiments are given in Figure 8B. When the P_{rbcS}-GAL4:PRO fusion was coelectroporated with the reporter plasmids containing ten GAL4 DNA binding sites in either orientation, luciferase activity increased 7- to 8-fold; no increase in activity was observed with either reporter plasmid when the constructs P_{rbcS}-GAL4:LZ or P_{rbcS}-GAL4 were coelectroporated. Coelectroporation with the activator construct pMA560, which has an acidic activation domain fused to the Gal4 DNA binding domain (Ma *et al.*, 1988), resulted in a 12-fold increase in activity. In contrast, no increase in luciferase activity was observed when any of these activators were coelectroporated with the reporter plasmid Δ89-LUC which lacks the GAL4 DNA binding site, although its basal level of activity is greater than that of either of the other reporters. Our results clearly demonstrated that the N-terminal proline-rich domain of GBF1 can stimulate transcription in plant protoplasts independently of its own DNA binding domain.

The proline-rich region of GBF1 stimulates transcription in mammalian cells

Recently it was suggested that unrelated eukaryotes may share common basic mechanisms for transcriptional activation (Ptashne, 1988). The yeast transcription factor, GAL4, for example, was shown to function in cells derived from various organisms, including plants (Fischer *et al.*, 1988; Kakidani and Ptashne, 1988; Ma *et al.*, 1988; Webster *et al.*, 1988). Furthermore, the tobacco TGACGT/C binding protein TGA1a was demonstrated to activate transcription in both human and plant *in vitro* transcription systems (Katagiri *et al.*, 1990; Yamazaki *et al.*, 1990). Both proteins, GAL4 and TGA1a, contain an acidic activation domain, whereas TGA1a is characterized by an additional glutamine-rich region (Katagiri *et al.*, 1989; Ma and Ptashne, 1987).

In order to investigate whether the proline-rich region of GBF1 would also function as a transcriptional activation domain in a heterologous system, we expressed different GAL4:GBF1 fusion proteins in mouse NIH3T3 cells together with a reporter plasmid. The GAL4:GBF1 fusion proteins consisted of the DNA binding domain of the yeast transcription factor GAL4 ($Gal4_{1-147}$) and various portions of GBF1 (Figure 9A). Two different reporter constructs were cotransfected: one contained the E1b TATA box linked to the CAT gene, the other differed by the addition of five GAL4 DNA binding sites linked to the E1b TATA box. The results of two independent experiments are summarized in Figure 9A and one representative CAT assay is shown in Figure 9B. No CAT activity was observed with the reporter plasmids alone (Figure 9B, lanes 1 and 7) or when the GAL4 DNA binding domain itself was coexpressed with either one of the reporter plasmids (lanes 2 and 8). A slight increase in CAT activity (~2-fold) was observed when a GAL4:GBF1 fusion protein (GAL4:GBF1ΔS, Figure 9A) containing the N-terminal domain of GBF1 (amino acids 1–222) was cotransfected with the reporter plasmid containing the GAL4 DNA binding sites (Figure 9B, lane 9). A further C-terminal deletion of the GBF1 coding region gave rise to the construct GAL4:Pro, which contained only sequences corresponding to the proline-rich region (amino acids 1–110, Figure 9A) fused to the GAL4 DNA binding domain. This deletion clone resulted in a protein which was even more active than GAL:GBF1ΔS (7- to 8-fold increase) when cotransfected with the reporter plasmid containing the

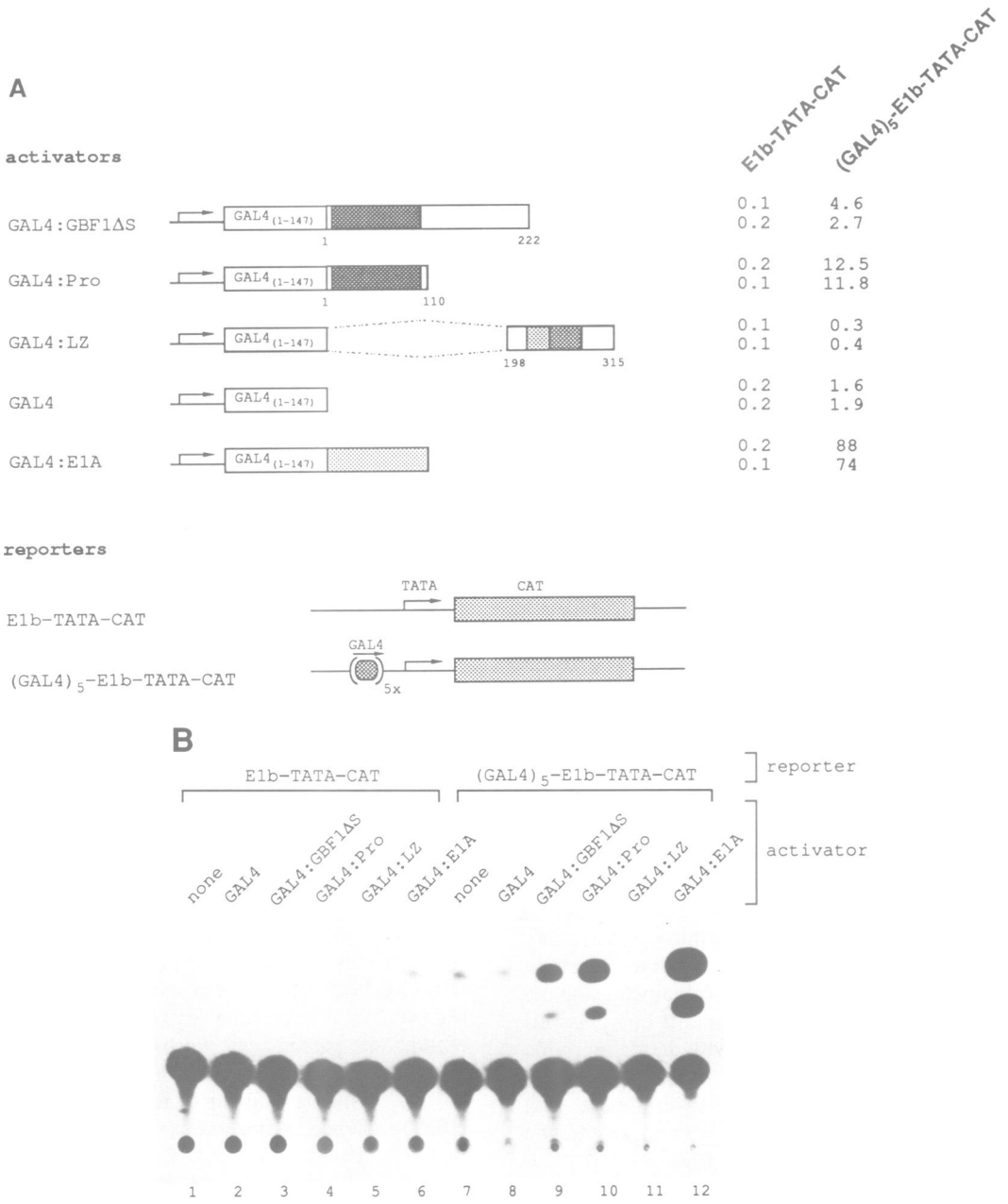


Fig. 9. The proline-rich region stimulates transcription in mammalian cells. (**A**) The activator constructs were composed of the GAL4 DNA binding domain [$GAL4_{(1-147)}$, Lillie and Green, 1989] and various parts of GBF1, driven by the SV40 promoter. Numbers below the boxes denote amino acids corresponding to GBF1. The functional domains of GBF1 are indicated and described in the legend to Figure 8. GAL4:E1A contains the activation domain of the viral transcription factor E-1A fused to the GAL4 DNA binding domain (Lillie and Green, 1989). The proteins were cotransfected with reporter plasmids (bottom) containing the CAT reporter gene, driven by the E1b TATA box; the two reporter constructs differ only in the presence or absence of five GAL4 DNA binding sites (Lillie and Green, 1989). Values on the right reflect CAT activity of cells harvested two days after cotransfection, the results of two independent experiments are shown. CAT values are expressed in arbitrary units relative to β-galactosidase activities. (**B**) Representative CAT assay. NIH3T3 cells were transfected with various activator and reporter constructs illustrated in panel A and indicated above each lane.

GAL4 DNA binding sites (Figure 9B, lane 10). In contrast, no activation was obtained with the construct GAL4:LZ (Figure 9A) where the C-terminal part of GBF1 (amino acids 198–315), encompassing the basic region and the leucine zipper, was fused to the GAL4 DNA binding domain (Figure 9B, lane 11). No transcriptional stimulation with any of these activator constructs was seen when a reporter plasmid lacking the GAL4 binding site was cotransfected (lanes 2–6). In summary, these results demonstrated that, when linked to the GAL4 DNA binding domain, the proline-rich region of GBF1 can function as a transcriptional activation domain in a heterologous system.

Discussion

Determination of high affinity GBF1 binding sites

Recently we identified three *Arabidopsis* cDNAs encoding proteins that interact with the G-box present in photoregulated and root specific promoters (Schindler *et al.*, 1992). Related G-box-like motifs, differing by one or two

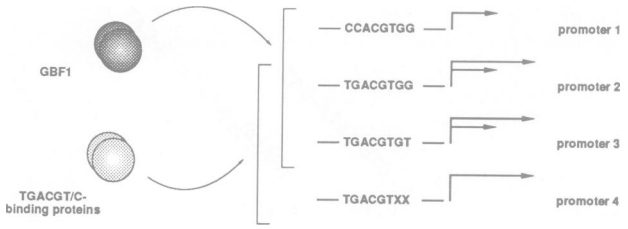


Fig. 10. Competition between GBF1 and a putative *Arabidopsis* TGACGT/C binding protein for promoter elements. Promoter 1 contains a G-box motif, promoter 2 contains the tgACGTGG motif (group IV), promoter 3 the tgACGTGt motif (group V) and promoter 4 a tgACGTXX motif, where XX represents any dinucleotide combination except GG or GT. In this model, promoter 1 will only be regulated by GBF1 (or GBF2 or GBF3) and promoter 4 only by the putative *Arabidopsis* TGACGT/C binding protein. The activity of promoters 2 and 3 will depend on the relative concentration of GBF1 and the TGACGT/C binding protein. Promoter activities would be expected to show both qualitative and quantitative differences depending on the bound transcriptional activator; these promoter differences are indicated schematically by large and small arrows.

base pairs, are also present in other plant promoters and appear to be required for their expression.

As a start towards discriminating between the three G-box binding proteins in *Arabidopsis* and to determine which promoters may be regulated by GBF1, we identified the exact DNA binding site requirements for GBF1. The random binding site selection assay provides a convenient and rapid method to investigate DNA sequences bound by a particular protein (Blackwell and Weintraub, 1990; Ekker *et al.*, 1991; Sun and Baltimore, 1991; Thiesen and Bach, 1990). While performing the assay the stringency of the selection can be monitored and high or low affinity binding sites can be selected. Furthermore, in contrast to binding studies involving oligonucleotides with defined sequences, this assay conveniently allows the determination of a large variety of different DNA binding sites. The results obtained using the random binding site selection assay support our previous findings, the GBF1 strongly interacts with the palindromic G-box, CCACGTGG (−4 to +4) motif. Furthermore, seven additional groups of DNA binding motifs were bound with similar high affinities by GBF1. All of them shared the ACGTG (−2 to +3) motif. In addition, our studies demonstrated that only certain base pair permutations at positions −4 and −3 reconstitute high affinity binding sites for GBF1. Motifs containing the combinations CC, AC, TG, CT and TT were bound with high affinity. In contrast, other dinucleotides—for example GT (G-rab) or TC (FSU-1A, Figure 7A and C)—were recognized with very low affinity. Furthermore, all high affinity binding sites contained either a G or a T residue at position +4 and thus sequences like the USF binding site (CCACGTGA) were not identified in the random binding site selection assay. It was surprising to find the relatively low representation of randomly selected sequences containing the palindromic CCACGTGG sequence, i.e. sequences with high G/C content. This may represent the fact that dCTP was used to label the PCR products and hence was included in the reactions at relatively low concentrations in comparison to other nucleotides. This interpretation is supported by the observation that the majority of the selected flanking sequences are observed to be rich at AT residues (Figure 5). However, if there was a bias in the selection it apparently was not so severe as to exclude any additional core sequences erroneously. We conclude this because the group IV core sequences (tgACGTGG), which contain only one less C residue than the palindromic group I sequences, were the most abundant group of selected sequences. Furthermore, when sequences that were missing from the random selection—such as the USF or G-chs-54 sequence (Figure 7)—were independently tested, they were shown not to bind. We note that the relative abundance of the different selected group sequences should not be interpreted in terms of reflecting different binding affinities of individual sequences; indeed, binding studies (data not shown) employing the sequences illustrated in Figure 5 and competition experiments (Figure 7B) demonstrated that these randomly selected motifs were bound with similar high affinities.

GBF1 binds with high affinity to the wheat histone 3 hexamer sequence

Previously we noticed that the DNA binding domain identified in GBF1 is very similar to the one found in the wheat protein HBP-1a (Schindler *et al.*, 1992). This similarity includes also the five amino acids flanking the basic residues of BR-A which we demonstrated were required for binding of GBF1 to the G-box. HBP-1a was shown to bind with high affinity to the wheat histone 3 hexamer sequence but with only very low affinity to the parsley *chs* G-box or the *cab-E* G-box-like element (Tabata *et al.*, 1991). In this present study our results suggest that GBF1 has a DNA binding affinity that is distinct in the following manner from that reported for HBP-1a. In contrast to HBP-1a, GBF1 interacted equally efficiently with the G-box (CCACGTGG) of the parsley *chs* promoter as with the tgACGTGG motif of the wheat histone 3 promoter. Furthermore, GBF1 also recognized the G-box-like motif (agACGTGG) of the *N.plumbaginifolia cab-E* promoter, although with somewhat lower affinity. These data suggest that although GBF1 and HBP-1a differ by only two amino acids within the basic region, the two proteins exhibit slightly different DNA binding site preferences. Furthermore, we were unable to alter the DNA binding properties of GBF1 by converting its DNA binding domain to the corresponding region of HBP-1a. Hence, it appears that structures outside the basic region contribute to the DNA binding affinity of these proteins; a finding in agreement with studies involving the DNA binding proteins Jun and Fos (Ransone *et al.*, 1990). Somewhat irrespective of the relative degree of binding of GBF1 and HBP-1a to the wheat histone 3 hexamer sequence, it is clear that in both cases these homodimeric proteins bind surprisingly strongly to an asymmetric binding site. The structural basis of this strong binding is not presently understood.

TGACGTXX motifs are recognized by GBF1 but only if the two 3′ base pairs are GG or GT

Using the random binding site selection assay we identified two groups of GBF1 binding sites containing the TGACGT motif. Group IV is characterized by tgACGTGG and resembles the wheat histone 3 hexamer sequence, whereas group V carries the sequence tgACGTGt. In addition to the wheat histone 3 promoter, TGACGT motifs were also identified within the CaMV 35S promoter (as-1 element) and the nopaline synthase (nos) and octopine synthase promoters (ocs). These DNA elements are bound by another class of plant bZIP proteins (TGACGT/C binding proteins) (Katagiri

et al., 1989; Singh *et al.*, 1990; Tabata *et al.*, 1991). The observed binding of GBF1 to group IV and V sequences and the lack of binding to the as-1 element or a mutant wheat histone 3 hexamer sequence (tgACGTtt) indicated that binding of GBF1 is strongly affected by the nature of the two base pairs following the TGACGT motif. We interpret the combined data from this study and from the study of others to indicate that promoters containing group IV or V sequences may be regulated either by GBF1 or by a TGACGT/C binding protein (Figure 10). Although no *Arabidopsis* genes for a TGACGT/C binding protein have been cloned so far, the finding of such proteins in many other plant species strongly suggests the existence of *Arabidopsis* homologues (Katagiri *et al.*, 1989; Singh *et al.*, 1990; Tabata *et al.*, 1991). Hence, as discussed later, the activity of a promoter containing group IV or V sequences (promoters 2 and 3, Figure 10) could depend on the relative concentrations of these two types of plant bZIP proteins. We have established the existence of two additional *Arabidopsis* GBF proteins (Schindler *et al.*, 1992) which also exhibit some affinity for group IV and V sequences (U.Schindler, A.E.Menkens and A.R.Cashmore, unpublished data) and we entertain the prospect of multiple *Arabidopsis* TGACGT/C binding proteins, such as identified in maize (Singh *et al.*, 1990). This situation would lead to a network of interactions resulting in differential gene expression.

Multiple plant bZIP proteins interact with DNA motifs containing the tetranucleotide ACGT

The G-box (CCACGTGG), bound by GBF proteins, and the TGACGT motifs, bound by TGACGT/C binding proteins, share the ACGT core sequence. Both protein classes belong to the family of bZIP proteins. This observation raises the question of whether all plant bZIP proteins will recognize the tetranucleotide ACGT. Based on the data revealed by the random binding site selection assay, we know that GBF1 requires the ACGT sequence; however, all high affinity binding sites are further characterized by a G residue (+3) flanking these four base pairs and only certain base pair combinations were allowed at positions −4, −3 and +4. These data suggest that individual plant bZIP proteins will probably exhibit different perferences concerning the base pairs flanking the ACGT motif. Consistent with this hypothesis is the observation that GBF1 binds with very low affinity to some naturally occurring G-box-like motifs which still contain the ACGT sequence. Since most of these elements were shown to be bound by nuclear proteins (Staiger *et al.*, 1989) it is conceivable that these proteins have DNA binding properties slightly distinct from GBF1. This situation is comparable to the binding behavior of the mammalian helix−loop−helix (HLH) proteins which are structurally similar to the bZIP proteins. Although all HLH proteins appear to bind to CANNTG sequences, individual proteins have different preferences concerning the internal two base pairs NN (Blackwell and Weintraub, 1990; Sun and Baltimore, 1991).

For GBF1 we clearly established a requirement for the ACGT core motif; however, the TGACGT/C binding proteins also interact with the TGACGC motif of the CaMV 35S promoter, indicating that binding of these proteins does not necessarily require the integrity of the ACGT tetranucleotide. Furthermore, the maize bZIP protein Opaque-2 binds to several DNA elements which lack the ACGT sequence (Lohmer *et al.*, 1991). The conclusion apparently is that most of the characterized plant bZIP proteins recognize, but do not necessarily demand, the core ACGT sequence. Finally, it is noteworthy that the apparent abundance of plant ACGT binding proteins may well reflect the intensive effort that has been devoted to screening for such proteins.

The proline-rich region of GBF1 activates transcription in a homologous and heterologous system

The ability of proline-rich regions to function as transcriptional activation domains has been described previously for the mammalian proteins CTF-1 (Mermod *et al.*, 1989) and AP-2 (Williams and Tjian, 1991). To determine whether GBF1 would function as a transcriptional activator in plant protoplasts, we investigated whether the full-length protein would stimulate transcription of a truncated CaMV 35S promoter containing multiple copies of the G-box sequence. Using this reporter construct we obtained high amounts of activity even in the absence of any activator plasmid. This endogenous activity was not increased by elevated expression of GBF1 and was most likely mediated by proteins related to GBF1 and present in the soybean protoplasts. As an alternative approach to investigate whether the proline-rich region could function as a transcriptional activation domain, we carried out experiments using the DNA binding site of the yeast transcription factor GAL4, since no endogenous GAL4 activity has been observed in plant protoplasts (Ma *et al.*, 1988). Our studies employing various fusion proteins between the GAL4 DNA binding domain and distinct parts of GBF1 indicated that the N-terminal proline-rich region had the potential to activate transcription. It seems likely that this proline-rich sequence performs a similar role in the context of GBF1.

The basic transcriptional mechanisms are similar in distantly related eukaryotes (Prashne, 1988). The plant transcription factor TGA1a, which has an acidic and glutamine-rich region, has been shown to function in a mammalian *in vitro* transcription system (Katagiri *et al.*, 1990; Yamazaki *et al.*, 1990). Since the proline-rich region of GBF1 activated transcription in plants, we sought to explore whether this region also augments transcription in other eukaryotes. We demonstrate that the proline-rich region of GBF1 is capable of stimulating transcription in mouse NIH3T3 cells when fused to the GAL4 DNA binding domain. These results are consistent with the observation that proline-rich regions function as transcriptional activation domains in mammalian cells (Mermod *et al.*, 1989; Williams and Tjian, 1991).

Whereas most G-box binding proteins—like GBF1—are characterized by proline-rich regions, the TGACGT/C binding proteins contain acidic or glutamine-rich regions (Katagiri *et al.*, 1989; Singh *et al.*, 1990; Tabata *et al.*, 1989, 1991; Weisshaar *et al.*, 1991). In mammalian systems it has been observed that proline-rich regions represent less potent transcriptional activation domains than the other two types (for review see Mitchell and Tjian, 1989) and we have made similar observations with the soybean protoplast system (Figure 8B). It would not be surprising if, in addition to these quantitative differences, the different transcriptional activators also exhibited qualitative differences. Given this likelihood, it is significant that we have demonstrated a clear overlap in the DNA binding properties of G-box binding

proteins and TGACGT/C binding proteins. Whereas protein−protein interactions are also likely to affect the formation of transcriptional complexes, in certain cases this overlap in binding properties would be expected to result in competition for the same DNA binding site. Thus, in addition to the relative cellular concentrations of GBF proteins affecting gene expression (Schindler *et al.*, 1992), we envisage the concentration of GBF1—and those of GBF2 and GBF3—relative to the predicted *Arabidopsis* TGACGT/C binding proteins to have an impact on the activity of certain promoters. We illustrate this in Figure 10 where the properties of hypothetical promoters (promoters 2 and 3), containing group IV and V sequences, are seen to be affected by the relative levels of GBF1 and TGACGT/C binding proteins. A full appreciation of the competing DNA binding interactions such as those depicted in Figure 10 will require a detailed characterization of all *Arabidopsis* G-box binding and TGACGT/C binding proteins, including a determination of the cellular distribution of these proteins and, as we have described in this report for GBF1, a detailed study of their DNA binding properties.

Materials and methods

Oligonucleotides

All oligonucleotides were synthesized on an Applied BioSystems DNA synthesizer 380B. Oligonucleotides of defined DNA sequence that were used for DNA binding studies were synthesized with *Bgl*II and *Bam*HI termini and cloned into the *Bgl*II and *Bam*HI sites of pBgl (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 templates GBF1(1−315), (199−288), (199−253), (219−288), (221−288), (219_{AA}−288) and GBF1-KL were generated by polymerase chain reaction (PCR) as described previously (Schindler *et al.*, 1992). In the case of GBF1(219_{AA}) and GBF1-KL the 5′ primers carried base pair substitutions which introduced alanines at positions 219 and 220 of GBF1 [GBF1(219_{AA})] or a lysine and a leucine at positions 225 and 230, respectively (GBF1-KL). The DNA sequence of the employed primers is available upon request. The *cab-E* promoter fragment A14 extending from −301 to −186 was described previously (Schindler and Cashmore, 1990). The GAL4 fusion constructs are based on the plasmid $pGAL4_{(1-147)}$ (Lillie and Green, 1989). GAL4:GBF1 was generated by ligating the *Bgl*II−*Kpn*I insert of SNG1 into the *Bam*HI and *Kpn*I site of $pGAL4_{(1-147)}$. The insert of SNG1 was originally generated by PCR, using two gene-specific primers, which introduced a *Bgl*II and an *Eco*RV site at the 5′ and 3′ ends of the coding region respectively. GAL4:GBF1ΔS was constructed by digestion of GAL4:GBF1 with *Sac*I and religation. GAL4:Pro was generated by ligating the blunt-ended *Hin*dIII fragment of pGBF1 (amino acids 1−110, Schindler *et al.*, 1992) into the *Sma*I site of $pGAL4_{(1-147)}$. GAL4:LZ was constructed by inserting the *Eco*RV fragment of p50 into the *Sma*I site of $pGAL4_{(1-147)}$. p50 contains the *Hin*dIII fragment of pGBF1 (amino acids 198−315) cloned into the *Hin*dIII site of pBluescript-SK(+). $(Gal4)_5$-E1b-TATA-CAT and E1b-TATA-CAT have been described previously (Lillie and Green, 1989). The *rbcS-1A* derivatives were made as follows. An *Eco*RI fragment containing the complete *rbcS-1A* promoter from −1700 to +21 fused to the GUS reporter gene was moved from prbcS-1A-GUS (Donald and Cashmore, 1990) into the *Eco*RI site of pGem4, generating prbcS-GPG. P_{rbcs}-GAL4 and P_{rbcS}-GAL4:Pro were made by digesting $pGAL4_{(1-147)}$ and GAL4:Pro with *Hin*dIII, filling with Klenow, then cutting with *Sac*I. The inserts were then ligated into prbc-GPG cut with *Sma*I and *Sac*I. P_{rbcS}-GAL4:LZ was made in two steps. First pBglG1, containing the same PCR fragment as SNG1 cloned into the *Bgl*II and *Eco*RV sites of pBgl (Donald *et al.*, 1990), was cut with *Spe*I, filled with Klenow, then cut with *Bgl*II. The insert was ligated into prbcS-GPG cut with *Sac*I, blunted with Klenow exonuclease, then cut with *Bam*HI, generating prbcS-GBF. The insert from P_{rbcs}-GAL4 cut with *Pst*I and *Sma*I was then ligated into prbcS-GBF cut with the same enzymes to generate P_{rbcS}-GAL4:LZ. Δ89-LUC was made in three steps. First pJD220 (a gift from Dr J.DeWet by way of Dr J.Ecker) was cut with *Sac*I and *Pst*I, and the insert was ligated into prbcS-GPG cut with the same enzymes, giving pLUC-PG. Next pLUC-PG was cut with *Pst*I and *Eco*RI, and the insert was ligated into pBluescript-SK(−) (Stratagene) cut with the same enzymes, yielding pLUC-BS. Finally, the 90 bp insert from pJD220 cut with *Eco*RV and *Pst*I was ligated into pLUC-BS cut with *Sma*I and *Pst*I, giving Δ89-LUC. $(GAL4)_{10}$-Δ89-LUC and $(GAL4)_{10r}$-Δ89-LUC were made by first ligating the 200 bp insert containing ten GAL4 binding sites from pMA558 (Ma *et al.*, 1988) cut with *Bam*HI and *Eco*RI into pGem4 cut with the same enzymes, giving pGAL4-PG. This was then cut with *Eco*RI and *Hin*dIII, and the insert was ligated into pBluescript-SK(−) cut with the same enzymes, yielding pGAL4-BS. Finally, pGAL4-BS was cut with *Bam*HI, and the insert was ligated into pΔ89-LUC cut with *Bam*HI, yielding plasmids with the GAL4 binding sites inserted in both orientations. The hybrid construct pHis-G1, expressing the fusion protein bGBF1, contains the *Bgl*II−*Bam*HI fragment of pBglG1, introduced into the *Bgl*II and *Bam*HI sites of pDS-MCS. pDS-MCS is a derivative of pDS56-6xHis (Abate *et al.*, 1990), carrying a multiple cloning site downstream of the six histidine residues (H.Beckmann, unpublished data).

Overexpression and purification of bGBF1

The fusion protein bGBF1 was overexpressed and purified on a Ni-NTA-agarose column (Qiagen) as described by Abate *et al.* (1990).

In vitro generation of proteins and DNA binding studies

In vitro transcription and *in vitro* translations, mobility shift assays and methylation interference experiments were performed as described previously (Schindler and Cashmore, 1990; Schindler *et al.*, 1992). Binding reactions for mobility shift assays contained 10^4 c.p.m. (10 fmol) of radiolabeled probe.

Random binding site selection

To select for GBF1 binding sites, a mixture of 64 base oligonucleotides were synthesized. The center of these oligonucleotides was composed of 14 random bases (see legend to Figure 5). Double strands were generated by extending from a primer (20 bases) which was annealed to the 3′ end of the 64 base oligonucleotide using DNA polymerase (Klenow) and [α-^{32}P]dATP. The double-stranded DNA pool was incubated with bGBF1 as described for the mobility shift assays. Protein−DNA complexes were separated on a 5% low ionic strength polyacrylamide gel (Donald *et al.*, 1990). The wet gel was autoradiographed, the DNA of the more slowly migrating protein−DNA complexes was eluted in elution buffer (0.5 M ammonium acetate, 1 mM EDTA) and recovered by ethanol precipitation. The DNA was amplified by PCR employing two primers complementary to the 3′ end of the positive or negative strand. [α-^{32}P]dCTP was incorporated into the PCR product during the reaction. The resulting DNA pool was subjected again to the same cycle; the cycle was repeated a total of five times. During each cycle the stringency of the protein−DNA binding conditions were increased, e.g. the protein concentration was decreased (from 500 to 50 ng renatured bGBF1) and the poly(dI−dC) concentrations were increased (from 0 to 5 μg). The PCR products obtained after the fifth cycle were digested with *Hin*dIII and *Bam*HI and ligated into the *Bam*HI and *Hin*dIII sites of pBluescript-SK(+). The inserts of positive clones were excised, radiolabeled with Klenow and [α-^{32}P]dATP and subjected to mobility shift assays using *in vitro* generated GBF1. Clones which were bound with the highest affinity were sequenced.

Transient assays in plant cells

SB-M (photomixotrophic) soybean cell cultures (a gift from Dr J.M.Widholm) were grown photomixotrophically as shaking batch cultures in KN1 medium (Rogers *et al.*, 1987). They were kept under constant light (100−150 μE/m^2/s) at 25°C on a gyratory shaker at 120 r.p.m. Cells harvested by centrifugation for 5 min at 100 *g* were resuspended in KN3M medium (KN3 medium with 400 mM mannitol; Rogers *et al.*, 1987) supplemented with 0.5% (w/v) Cellulysin and 0.1% (w/v) Macerase. After shaking at 40 r.p.m. overnight at 25°C in the dark on a gyratory shaker, the suspension was filtered through 250 μm and 53 μm mesh nylon screens, then protoplasts were harvested by centrifugation for 5 min at 100 *g*. The pellet was resuspended in 30% Percoll in KN3M medium, then overlaid with KN3M medium and centrifuged for 5 min at 100 *g*. Protoplasts banding at the interface were collected, diluted 5-fold with KN3M medium, then harvested by centrifugation for 5 min at 100 *g*. The resulting pellet was resuspended in KN3M at ~10^6 cells/ml, transferred to a 100 mM Petri dish and cultured for 2 days at low light (20 μE/m^2/s) at 25°C. Protoplasts were electroporated as described by Lin *et al.* (1987) with the following modifications. Protoplasts were transferred to 15 ml tubes and heat-shocked for 5 min at 37°C. They were then mixed with plasmid, pNosCAT (as internal control for electroporation efficiency), and carrier salmon sperm DNA at final concentrations of 50, 50 and 100

μg/ml respectively, and left for 10 min at room temperature. Next they were diluted with one-third volume 30% (w/v) PEG 6000, 120 mM $MgCl_2$ in KN3M, and 0.8 ml aliquots were transferred to 4 mm Bio-Rad electroporation cuvettes and stored for 5 min on ice. They were then electroporated with the Bio-Rad apparatus set at 150 V, 960 μF, which gave time constants varying from 50 to 60 ms, and stored for 10 min on ice. The cells were then transferred to 10 ml KN3M, and harvested by centrifugation for 5 min at 100 *g*. They were resuspended in 1 ml KN3M, transferred to 60 mm dishes and cultured for 2 days in low light (20 $\mu E/m^2/s$) at 25°C. Cells were then transferred to 15 ml tubes and harvested by centrifugation for 5 min at 100 *g*. They were resuspended in 150 μl 2 × extraction buffer (200 mM KPO_4 pH 7.8, 2 mM DTT), transferred to microfuge tubes and lysed by sonication. The extract was spun for 10 min in a microfuge and the supernatant was used for luciferase assays. 50 μl aliquots were placed in disposable luminometer cuvettes, mixed with 60 μl 5 × assay buffer (500 mM KPO_4) pH 7.8, 50 mM $MgCl_2$, 25 mM ATP, 5 mM DTT), and diluted to 300 μl with distilled water. The cuvette was then placed in an Analytical Luminescence Monolight 2001 luminometer, the reaction was initiated by injecting 100 μl 1 mM luciferin, and the peak during the initial 10 s was measured. Results were normalized to relative transformation efficiencies by CAT activity assayed in 50 μl aliquots of each extract as described (Sambrook *et al.*, 1989).

Transient expression in mammalian cells

Transfection of mouse NIH3T3 cells and CAT assays, normalized to relative transfection efficiencies by β-galactosidase expression, were carried out as described previously (Beckmann *et al.*, 1990). In each transfection, reporter plasmids (5 μg) were transfected along with the indicated activator plasmid (10 μg) plus 3.5 μg pCH110 (Lee *et al.*, 1984), expressing the bacterial β-galactosidase gene, and pUC19 DNA to bring the total amount of transfected DNA to 21 μg. Cells were harvested and assayed for CAT activity 2 days after transfection.

Acknowledgements

We thank J.Widholm for his gift of the SB-P cell line, and J.Ma for providing plasmids pMA558, pMA560 and pNOSCAT. Similarly, we thank M.Thomashow for providing us with the sequence of a G-box-like element from the promoter of *cor180*. We thank M.Ahmad for the synthesis of oligonucleotides and, together with J.Chan and A.Menkens, for providing critical comments concerning this manuscript. This work was supported by NIH grant GM 38409 to A.R.C. and by funds from the Howard Hughes Medical Institute to T.K. W.B.T. was supported by NSF postdoctoral fellowship DCB-8807578.

References

Abate,C., Luk,D., Gentz,R., Rauscher,R.I. and Curran,T. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1032–1036.
Beckmann,H., Su,L.-K. and Kadesch,T. (1990) *Genes Dev.*, **4**, 167–179.
Blackwell,T.K. and Weintraub,H. (1990) *Science*, **250**, 1105–1110.
Carey,M., Lin,Y.W., Green,M.R. and Ptashne,M. (1990) *Nature*, **345**, 361–364.
Castresana,C., Garcia-Luque,I., Alonso,E., Malik,V.S. and Cashmore,A.R. (1988) *EMBO J.*, **7**, 1929–1936.
Courey,A.J., Holtzman,D.A., Jackson,S.P. and Tjian,R. (1989) *Cell*, **59**, 827–836.
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.
Ekker,S.C., Young,K.E., von Kessler,D.P. and Beachy,P.A. (1991) *EMBO J.*, **10**, 1179–1186.
Fischer,J.A., Giniger,E., Maniatis,T. and Ptashne,M. (1988) *Nature*, **332**, 853–856.
Frankel,A.D. and Kim,P.S. (1991) *Cell*, **65**, 717–719.
Gerster,T., Balmaceda,C.G. and Roeder,R.G. (1990) *EMBO J.*, **9**, 1635–1643.
Gregor,P.D., Sawadogo,M. and Roeder,R.G. (1990) *Genes Dev.*, **4**, 1730–1740.
Gronemeyer,H. *et al.* (1987) *EMBO J.*, **6**, 3985–3994.
Guiltinan,M.J., Marcotte,W.J. and Quatrano,R.S. (1990) *Science*, **250**, 267–271.
Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885–894.
Hu,J.C., O'Shea,E.K., Kim,P.S. and Sauer,R.T. (1990) *Science*, **250**, 1400–1403.
Johnson,P.F. and McKnight,S.L. (1989) *Annu. Rev. Biochem.*, **58**, 799–839.
Jones,N.C., Rigby,P.W.J. and Ziff,E.B. (1988) *Genes Dev.*, **2**, 267–281.
Kakidani,H. and Ptashne,M. (1988) *Cell*, **52**, 161–167.
Katagiri,F., Lam,E. and Chua,N.-H. (1989) *Nature*, **340**, 727–730.
Katagiri,F., Yamazaki,K., Horikoshi,M., Roeder,R.G. and Chua,N.-H. (1990) *Genes Dev.*, **4**, 1899–1909.
Kumar,V., Green,S., Stack,G., Berry,M., Jin,J.-R. and Chambon,P. (1987) *Cell*, **51**, 941–951.
Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.
Latchman,D.S. (1990) *Biochem. J.*, **270**, 218–289.
Laughon,A. and Scott,M.P. (1984) *Nature*, **310**, 25–31.
Lee,F., Hall,C.V., Ringold,G.M., Dobson,D.E., Luh,J. and Jacob,P.E. (1984) *Nucleic Acids Res.*, **12**, 4191–4206.
Lillie,J.W. and Green,M.R. (1989) *Nature*, **338**, 39–44.
Lin,W., Odell,J.T. and Schreiner,R.M. (1987) *Plant Physiol.*, **84**, 856–861.
Lohmer,S., Maddaloni,M., Motto,M., Di Fonzo,N., Hartings,H., Salamini,F. and Thompson,R.D. (1991) *EMBO J.*, **10**, 617–624.
Ma,J. and Ptashne,M. (1987) *Cell*, **48**, 847–853.
Ma,J., Przibilla,E., Hu,J., Bogorad,L. and Ptashne,M. (1988) *Nature*, **334**, 631–633.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Mermod,N., O'Neill,E.A., Kelly,T.J. and Tjian,R. (1989) *Cell*, **58**, 741–753.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Müller,I.M., Schaffner,W. and Matthias,P. (1990) *EMBO J.*, **9**, 1625–1634.
Oeda,K., Salinas,J. and Chua,N.-H. (1991) *EMBO J.*, **10**, 1793–1802.
Ptashne,M. (1988) *Nature*, **335**, 683–689.
Ransone,L.J., Wamsley,P., Morley,K.L. and Verma,I.M. (1990) *Mol. Cell. Biol.*, **10**, 4565–4573.
Rasmussen,R., Benvegnu,D., O'Shea,E., Kim,P.S. and Alber,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 561–564.
Rogers,M.D., Ogren,W.L. and Widholm,J.M. (1987) *Plant Physiol.*, **84**, 1451–1456.
Rosenberg,P.J., Schroeder,C., Preiss,A., Kienlin,A., Cote,S., Riede,I. and Jäckle,H. (1986) *Nature*, **319**, 336–339.
Santoro,C., Mermod,N., Andrews,P.C. and Tjian,R. (1988) *Nature*, **334**, 218–334.
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.
Sawadogo,M. and Roeder,R.G. (1985) *Cell*, **43**, 165–176.
Schindler,U. and Cashmore,A.R. (1990) *EMBO J.*, **9**, 3415–3427.
Schindler,U., Menkens,A.E., Beckmann,H., Ecker,J.R. and Cashmore,A.R. (1992) *EMBO J.*, **11**, 1261–1273.
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.
Sun,X.-H. and Baltimore,D. (1991) *Cell*, **64**, 459–470.
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.
Thiesen,H.-J. and Bach,C. (1990) *Nucleic Acids Res.*, **18**, 3203–3209.
Webster,N., Jin,J.R., Green,S., Hollis,M. and Chambon,P. (1988) *Cell*, **52**, 169–178.
Weintraub,H., Davis,R., Tapscott,S., Thayer,M., Krause,M., Benezera,R., Blackwell,T.K., Turner,D., Rupp,R., Hollenberg,S., Zhuang,Y. and Lassar,A. (1991) *Science*, **251**, 761–766.
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,T., Admon,A., Leuscher,B. and Tjian,R. (1988) *Genes Dev.*, **2**, 1557–1569.
Yamazaki,K.-I., Katagiri,F., Imaseki,H. and Chua,N.-H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 7035–7039.
Ziff,E.B. (1990) *Trends Genet.*, **6**, 69–72.

Received on October 18, 1991; revised on January 10, 1992