The EMBO Journal vol.9 no.6 pp.1727 – 1735, 1990

The plant G box promoter sequence activates transcription in *Saccharomyces cerevisiae* and is bound *in vitro* by a yeast activity similar to GBF, the plant G box binding factor

Robert G.K.Donald, Ulrike Schindler, Alfred Batschauer and Anthony R.Cashmore

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

Communicated by M.van Montagu

G box and I box sequences of the *Arabidopsis thaliana* ribulose-bisphosphate-1,5-carboxylase small subunit (RBCS) promoter are required for expression mediated by the *Arabidopsis rbcS-1A* promoter in transgenic tobacco plants and are bound *in vitro* by factors from plant nuclear extracts termed GBF and GA-1, respectively. We show here that a −390 to −60 *rbcS-1A* promoter fragment containing the G box and two I boxes activates transcription from a truncated *iso*-1-cytochrome *c* (CYC1) gene promoter in *Saccharomyces cerevisiae*. Mutagenesis of either the *rbcS-1A* G box or both I box sequences eliminated the expression mediated by this fragment. When polymerized, I box oligonucleotides were also capable of enhancing expression from the truncated CYC1 promoter. Single-copy G box sequences from the *Arabidopsis rbcS-1A*, *Arabidopsis Adh* and tomato *rbcS-3A* promoters were more potent activators and were used in mobility shift assays to identify a DNA binding activity in yeast functionally similar to GBF. In methylation interference experiments, the binding specificity of the yeast protein was indistinguishable from that obtained with plant nuclear extracts.

Key words: GBF/G box/heterologous UAS/yeast transcription

Introduction

Many of the features of the transcriptional machinery of yeast and higher eukaryotes are functionally related and evolutionarily conserved. Conservation of function and structure has been demonstrated for several yeast and animal proteins which bind to promoter sequences and activate transcription (Struhl, 1987; Buratowski *et al.*, 1988; Cavallini *et al.*, 1988; Chodosh *et al.*, 1988; Harshman *et al.*, 1988; Hayes *et al.*, 1988; Horikoshi *et al.*, 1988; Jones *et al.*, 1988; Lin *et al.*, 1988, 1989). The observation that the yeast Gal4 transcription factor can activate expression in *Drosophila* cells (Fisher *et al.*, 1988), in mammalian cells (Kakidani and Ptashne, 1988; Webster *et al.*, 1988) and in tobacco cells (Ma *et al.*, 1988) when provided with its cognate binding site, suggests a measure of conservation for the mechanisms of transcription in plants and other eukaryotes. However, relatively little is known about the details of transcriptional activation of plant promoters. Evidence for the functional significance of some sequence-specific DNA binding proteins in transcriptional activation of plant gene expression has been provided by the apparent requirement of cognate DNA sequences for promoter expression (Kuhlemeier *et al.*, 1987, 1988; Lam *et al.*, 1989; Donald and Cashmore, 1990).

We have been studying the regulation of *rbcS-1A* promoter expression in transgenic tobacco. We have shown that mutation of sequences recognized by DNA binding proteins GBF and GA-1 within the *rbcS-1A* promoter, profoundly affect expression of β-glucuronidase (GUS) or alcohol dehydrogenase (*Adh*) reporter genes (Donald and Cashmore, 1990; U.Schindler and A.R.Cashmore, submitted).

The G box binding factor (GBF), identified from *Arabidopsis* and tomato nuclear extracts, binds specifically *in vitro* to a sequence (C/A-CACGTGGCA) termed the G box which is strongly conserved among the RBCS gene family (Giuliano *et al.*, 1988; Manzara and Gruissem, 1988). A G box has also been shown to be important for expression of the parsley chalcone synthase promoter in parsley cell cultures and related sequences are present in other chalcone synthase promoters (Schulze-Lefert *et al.*, 1988; Staiger *et al.*, 1989). A G box is also present in the *Arabidopsis Adh* gene (Chang and Meyerowitz, 1986). Experiments with *Arabidopsis* cell cultures have shown that this *Adh* G box is bound *in vivo* by a protein factor (Ferl and Laughner, 1989).

Factor GA-1 interacts with the conserved GATA box region located immediately proximal to the TATA box of many chlorophyll-a/b binding protein (CAB) gene promoters (Castresana *et al.*, 1987; Gidoni *et al.*, 1989; U.Schindler and A.R.Cashmore, submitted). The same protein also interacts with the conserved I boxes (Giuliano *et al.*, 1988) or sequences 2 (Manzara and Gruissem, 1988) of the *Arabidopsis rbcS-1A* promoter, which have as consensus a GATAAG motif and which are located further upstream in RBCS promoters (U.Schindler and A.R.Cashmore, submitted).

With the aim of identifying conserved elements common to both the plant and yeast transcriptional apparatus we made use of a yeast CYC1 – *lacZ* expression vector (Guarente and Ptashne, 1981; Guarente, 1983) to identify possible *rbcS-1A* sequences capable of activating promoter expression in yeast. Deletion of CYC1 upstream elements results in an inactive truncated CYC1 – *lacZ* gene fusion which can be used to test the ability of heterologous promoter fragments to re-activate CYC1 – *lacZ* expression (Guarente and Ptashne, 1981; Figure 1A). We show here that the sequence recognized *in vitro* by plant factor GBF, can act as an upstream activating sequence (UAS) in promoting expression from the truncated yeast CYC1 promoter in yeast and we identify a specific yeast DNA binding activity (yGBF) with properties similar to GBF.

Results

Identification of RBCS promoter sequences which can activate expression from the truncated CYC1 promoter in yeast

Fragments of the *Arabidopsis rbcS-1A* promoter were tested for their ability to activate expression from the truncated

CYC1 promoter. Fragment 3, consisting of 330 bp upstream of the −60 *Hin*dIII site, was able to restore 33% of the level of β-galactosidase activity observed for the full CYC1 promoter (Figure 1B). In contrast, larger *rbcS-1A* fragments spanning 1.6 and 1.1 kb of promoter upstream of the −60 bp *Hin*dIII site (fragments 1 and 2) were incapable of enhancing the basal expression of the truncated CYC1 promoter. Apparently sequences upstream of −390 have a negative effect on the expression conferred by downstream *rbcS-1A* sequences on the truncated CYC1 promoter.

Conserved DNA sequence motifs corresponding to sites of specific binding by plant nuclear factors GA-1 and GBF reside within fragment 3 (Giuliano *et al.*, 1988; Donald and Cashmore, 1990; U.Schindler and A.R.Cashmore, in preparation). The 12 bp site-specific substitution mutations in recognition sites for either GBF or GA-1 drastically affect expression from the full 1.7 kb *rbcS-1A* promoter in transgenic tobacco (Donald and Cashmore, 1990). To examine the contribution of these conserved sequence elements to the expression mediated by fragment 3 in yeast, the same mutations tested *in planta* were tested in yeast (Figure 1B). Strikingly, both G box (fragment 4) and double I box (fragment 5) mutations reduced the expression mediated by fragment 3 to background levels.

Fragment 3 also contains sequences showing homology with binding sites for pea factor GT-1 (Figure 1B; Green *et al.*, 1987, 1989). In contrast with G and I box mutations, mutation of these sequences in the context of the full *rbcS-1A* promoter had no effect on reporter expression in transgenic plants (Donald and Cashmore, 1990). When tested in the context of fragment 3 expression in yeast, these mutations also had no effect (data not presented). Surprisingly, the effect of deletion Δ12 (fragment 7), which removes 89 bp including a putative GT-1 site and which had little or no effect *in planta*, increased the level of expression mediated by fragment 3 from 33 to 59% the level of expression observed for the wild-type CYC1 promoter. Since a similar enhancement is also observed with overlapping deletion Δ11 (fragment 6) which does not span any putative GT-1 sites, the effect is probably not due to the loss of a GT-1 binding site.

Fragments 9 and 10, which constitute the same fragment placed in either orientation with respect to the truncated CYC1 promoter, also show an enhancement of β-galactosidase expression when compared with the larger fragment 3. This fragment, termed the light regulatory element (LRE), can confer *in planta* light-regulated expression on a heterologous *Adh* gene (Donald and Cashmore, 1990).

The enhancement in β-galactosidase expression observed for deleted derivatives of fragment 3 (fragments 6, 7, 9 and 10) can be attributed either to the deletion of DNA sequences inhibitory to yeast expression, or to the effect of position of *rbcS-1A* sequences relative to the truncated yeast promoter. In either event, the G box is still required for the expression mediated by fragment 3 bearing the Δ12 deletion (compare fragments 7 and 8). The double I box mutation was not tested in the context of the Δ12 deletion.

The G box can activate transcription from the truncated CYC1 promoter in yeast

Having established the requirement for the G box and both I boxes for the expression mediated by *rbcS-1A* promoter

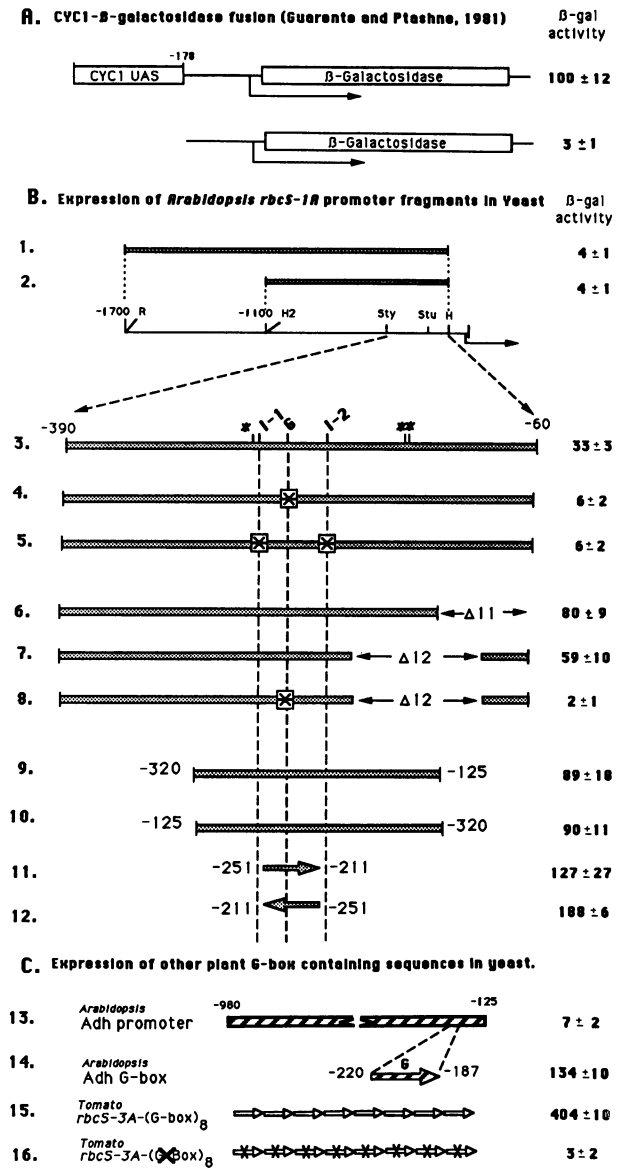

Fig. 1. Effect of *Arabidopsis rbcS-1A* and *Adh* promoter fragments on transcription in yeast cells. (**A**) CYC1 expression vectors (Guarente and Ptashne, 1981). Arrows indicate the direction of transcription, the number 178 the distance (bp) relative to the most upstream transcription initiation site (Guarente and Mason, 1983). The top construct (pLG669-Z) carries the intact CYC1 promoter. The lower construct (pLG670-Z) carries a truncated CYC1 promoter deleted of its UAS elements and is the basis of remaining constructs shown below. In these constructs, the indicated plant promoter sequences are fused to this truncated CYC1 promoter. (**B**) Fragments derived from the 1.7 kb *rbcS-1A* promoter. Restriction enzyme cleavage sites shown are *Eco*RI (R), *Hin*dII (H2), *Sty*I (Sty), *Stu*I (Stu) and *Hin*dIII (H). The location of the *rbcS-1A* promoter G box element, I box elements and corresponding 12 bp substitution mutations is indicated. An asterisk (*) marks the location of *rbcS-1A* sequences homologous to pea factor GT-1 recognition sites within the pea *rbcS-3A* promoter (Green, 1987, 1989). These putative *rbcS-1A* GT-1 sequences were also targeted by 12 bp substitutions. (**C**) Fragments derived from the *Arabidopsis Adh* and tomato *rbcS-3A* promoter. Arrows indicate orientation of *rbcS-1A* and *Adh* fragments relative to the direction of *rbcS-1A* and *Adh* gene transcription. Yeast a cells (146a; MATa *ura3 his4 leu2 trp1*) were transformed with each of these plasmids and β-galactosidase levels of at least three individual transformants, grown in selective minimal glucose media, were determined. Numbers represent the mean and standard error for these values. The values shown were reproducible in two separate experiments. Yeast α and a/α cells (Materials and methods) were also transformed with selected constructs, and no significant differences in relative β-galactosidase expression levels were observed.

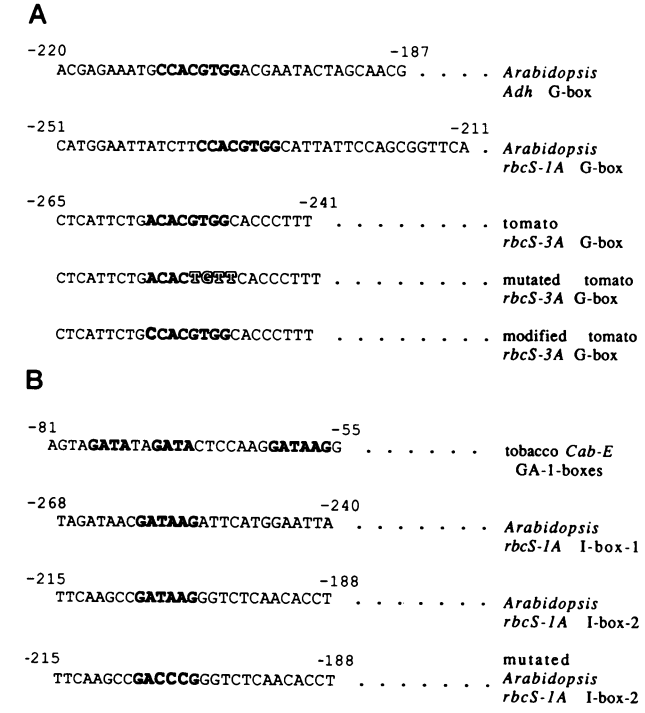

Fig. 2. Plant promoter sequences containing plant GBF (**A**) and GA-1 (**B**) DNA binding sites used in gel mobility-shift experiments or expression studies. The *Arabidopsis Adh* or *rbcS-1A* G boxes were subcloned from genomic sequences into pBS (Bluescript) and pUC19 plasmid vectors respectively; the tomato G-box sequences and all GA-1 sequences were chemically synthesized as oligonucleotides (see Materials and methods). Oligonucleotides were flanked by 6 bp recognition sites for *Bgl*II and *Bam*HI (not shown) to facilitate subcloning and subsequent polymerization. Nucleotides emphasized in bold text represent GBF or GA-1 binding motifs.

fragment 3 from the truncated CYC1 promoter, we sought to determine whether these sequences could function independently as activators.

Several DNA fragments containing G box motifs were tested for their ability to activate β-galactosidase expression from the truncated CYC1 promoter (Figures 1B and C and 2A). A fragment containing 40 bp of *rbcS-1A* promoter including the G box motif could activate β-galactosidase expression from the truncated CYC1 promoter, independent of orientation and to levels higher than the wild-type CYC1 promoter (fragments 11 and 12). This fragment was excised from a mutated *rbcS-1A* DNA template in which restriction endonuclease recognition sites were introduced substituting for and eliminating both *rbcS-1A* I box motifs (see Materials and methods; Donald and Cashmore, 1990). Similarly, a fragment containing 33 bp of the *Arabidopsis Adh* promoter, including a G box motif (Figure 2A), was also able to activate expression from the truncated CYC1 promoter (fragment 14). However, as was observed in the case of *rbcS-1A* fragments 1 and 2, a larger *Adh* promoter fragment (fragment 13) including the G box and spanning 755 bp of promoter upstream of −125 bp, was incapable of activating β-galactosidase expression.

The sequence specificity of the observed G-box-dependent activation of truncated CYC1 promoter was tested by comparing the expression mediated by the tomato *rbcS-3A* G box and a mutated derivative. Two 25 bp oligonucleotides were synthesized bearing the wild-type and mutant *rbcS-3A* G box motif, respectively (Figure 2A). The sequences were

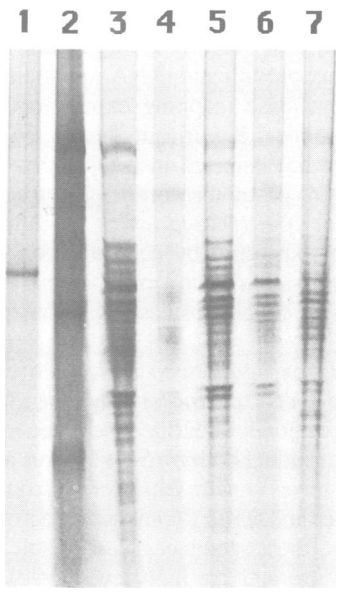

Fig. 3. RNase protection mapping of the 5′ ends of CYC1−*lacZ* mRNAs. Total RNA was prepared from yeast **a** cells transformed with the constructs indicated. RNA samples (5−10 μg) were hybridized with an SP6-generated riboprobe uniformly labeled with [α-^{35}S]UTP and digested with RNase A/T1 (Materials and methods). The probe is complementary at its 5′ end to the chimeric CYC1−*lacZ* mRNAs expressed by these constructs. **Lane 1**, a 172 bp RNA marker synthesized from a riboprobe SP6 positive control template (Promega). **Lane 2**, the 210 bp CYC1−*lacZ* probe used in the protection assay. **Lane 3**, pLG669-Z (CYC1−*lacZ*). **Lane 4**, pLG670-Z (truncated CYC1−*lacZ*). **Lanes 5−7**, constructs in which respectively the *Arabidopsis rbcS-1A* G box (Figure 1A, fragment 11), *Adh* G box (fragment 14), and *rbcS-1A* LRE (fragment 9) were fused to the truncated CYC1 promoter. Between six and eight prominent bands are common to lanes 3, 5, 6 and 7. Differences observed between these lanes in the relative intensity of less prominent bands are presumed to be due to incomplete RNase digestion.

polymerized 8 times in direct repeats and fused to the truncated CYC1 promoter for expression analysis in yeast. The polymerized G box oligonucleotides were used for expression analysis rather than the single G box fragments described above, since *in vitro* DNA binding experiments indicated a stronger binding capacity of plant GBF for oligomerized G box sequences (data not presented). Whereas the mutated polymerized tomato *rbcS-3A* G box sequence (fragment 16) was unable to activate the truncated CYC1 promoter, the wild-type polymerized sequence (fragment 15) gave 4 times the level of β-galactosidase activity observed for the wild-type CYC1 promoter.

To ensure that the G-box-dependent activation of truncated CYC1 promoter was reflected at the level of mRNA transcript, RNase protection experiments were performed. Total RNA was prepared from yeast cells transformed with CYC1−*lacZ* fusion constructs, hybridized with an SP6-polymerase-generated CYC1−*lacZ* anti-sense RNA probe and treated with RNase. Results shown in Figure 3 indicate that the same multiple sites of transcript initiation detected for the full CYC1 promoter (lane 3) were observed for *Arabidopsis rbcS-1A* and *Adh* G box constructs (Figure 1B, fragments 11 and 14; Figure 3, lanes 5 and 6). The same array of transcription start sites was also observed for cells expressing the *rbcS-1A* LRE construct (fragment 9) in which the G box is located further upstream relative to the truncated CYC1 promoter (lane 7). The number of transcripts we

observed and the position of the transcription start sites are similar to those described by Guarente and Mason (1983) who detected six prominent mRNA initiation sites spanning 34 bp in S1 nuclease mapping experiments of the same CYC1−*lacZ* gene. Significant levels of CYC1−*lacZ* transcript were not detected in cells transformed with the truncated CYC1 promoter control construct (lane 4). In this experiment, the relative levels of chimeric CYC1−*lacZ* transcript expressed by cells transformed with these constructs were approximately proportional to the observed relative levels of β-galactosidase (data not presented).

Polymerized I box sequences activate expression from the truncated CYC1 promoter in yeast

Several oligonucleotides bearing motifs recognized by plant factor GA-1, including both *rbcS-1A* I boxes and the tobacco *Cab-E* GATA boxes (Figure 2B), were fused to the truncated CYC1 promoter and tested for enhancing activity. In contrast with the G box, these sequences were unable, when present in single copy, to activate high levels of expression from the truncated CYC1 promoter. Low levels of expression, at 2- to 3-fold background levels, were detected for each fusion construct (data not shown). However, higher levels of expression were detected with polymerized *rbcS-1A* I box sequences. The sequence-specificity of this expression was investigated by comparing the expression mediated by a polymerized *rbcS-1A* I box 2 with a polymerized mutated derivative in which a 3 bp substitution had been introduced into the GATAAG motif (Figure 2B). In the experiment shown in Figure 4 the dimeric, tetrameric and octameric I box 2 sequence conferred a step-wise enhancement of CYC1−*lacZ* expression that was not observed in the case of the mutated derivatives. The I box 1 and I box 2 octamers gave 19 and 38%, respectively, the expression conferred by the full CYC1 promoter.

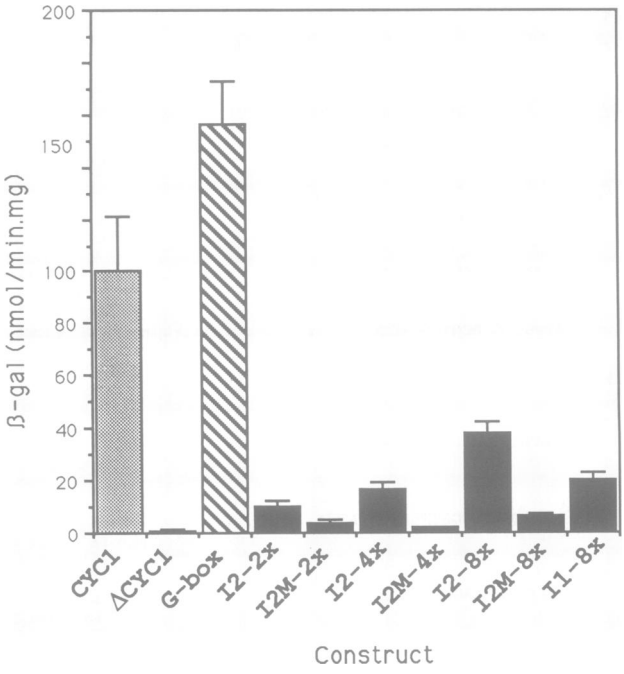

Fig. 4. A comparison of expression mediated by *rbcS-1A* G box and I box sequences fused to the truncated CYC1 promoter. The indicated *rbcS-1A* promoter elements were fused to the truncated CYC1 promoter and β-galactosidase reporter expression analyzed in yeast transformed with these constructs as described (legend to Figure 1 and Materials and methods). Control constructs (see also Figure 1): CYC1, full CYC1 promoter; Δ-CYC1, truncated CYC1 promoter. Test sequences fused to the truncated CYC1 promoter: *rbcS-1A* G box, −251 to −211 (see also Figure 1); I box 2 dimer (I2-2x); mutated I box 2 dimer (I2M-2x); I box 2 tetramer (I2-4x); mutated I box 2 tetramer (I2M-4x); I box 2 octamer (I2-8x); mutated I box 2 octamer (I2M-8x); I box 1 octamer (I1-8x). Small differences in relative expression of control constructs and the *rbcS-1A* G box construct compared with values shown in Figure 1 reflect the experimental variance of the separate sets of experiments. Error bars indicate the standard error of the mean of each population of independent yeast transformants analyzed.

A yeast GBF-like activity shows specificity similar to plant GBF in in vitro DNA binding assays

The activation of β-galactosidase reporter expression from the truncated CYC1 promoter in yeast by plant G box sequences suggested a recognition of these sequences by a yeast factor or factors. In an attempt to identify a yeast GBF homolog, yeast nuclear extracts were prepared and compared with *Arabidopsis* extracts for their ability to specifically bind G-box-containing DNA fragments in gel mobility-shift assays.

A radiolabeled *rbcS-1A* promoter fragment (−251 to −211; fragment 11) containing the G box motif was incubated with yeast or *Arabidopsis* nuclear extract in the presence of non-specific competitor DNA [poly(dI−dC)]. Protein-complexed and free DNA fragments were separated on a low ionic strength polyacrylamide gel. Two prominent closely-spaced slowly-migrating protein DNA complexes were observed when yeast nuclear extract was added (Figure 5, lane 2, upper panel). The addition of *Arabidopsis* nuclear extract generated a slightly faster migrating protein−DNA complex observed as a single broad band (Figure 5, lane 2, lower panel).

To determine the specificity of these protein−DNA interactions, binding reactions with both extracts were carried out in the presence of different competitor DNAs. As shown in Figure 5, a plasmid containing eight copies of the oligonucleotide representing the wild-type tomato *rbcS-3A* G box motif competes for the binding of the yeast activity as well as the *Arabidopsis* activity (lanes 3−5). A plasmid containing eight copies of the mutated DNA binding site (lanes 6−8) and the plasmid vector (lanes 9−11) alone do not show this effect (compare lanes 5, 8 and 11).

Competition is also observed for both nuclear extracts when the single tomato *rbcS-3A* G box oligonucleotide is included in the binding reaction (Figure 5, lanes 12−14). Similarly, the competition of a modified G box oligonucleotide ('G-boxpp'), characterized by a single base pair substitution (Figure 2A), was tested. The mutation converts the core binding motif as it is present within the tomato *rbcS-3A* promoter to a perfect palindrome (CCACGTGG) as it is found within the *Arabidopsis rbcS-1A* or *Adh* promoters. The addition of this oligonucleotide to the binding reaction also results in a strong competition (lanes 15−17) with both nuclear extracts but in the case of *Arabidopsis* extract, the modified oligonucleotide competes more efficiently (compare lanes 14 and 17). However, the stronger competition observed for the modified compared with the unmodified tomato G box oligonucleotides for the *Arabidopsis* GBF activity is not observed when tomato nuclear extract is used (data not shown).

To more specifically delineate the DNA sequences

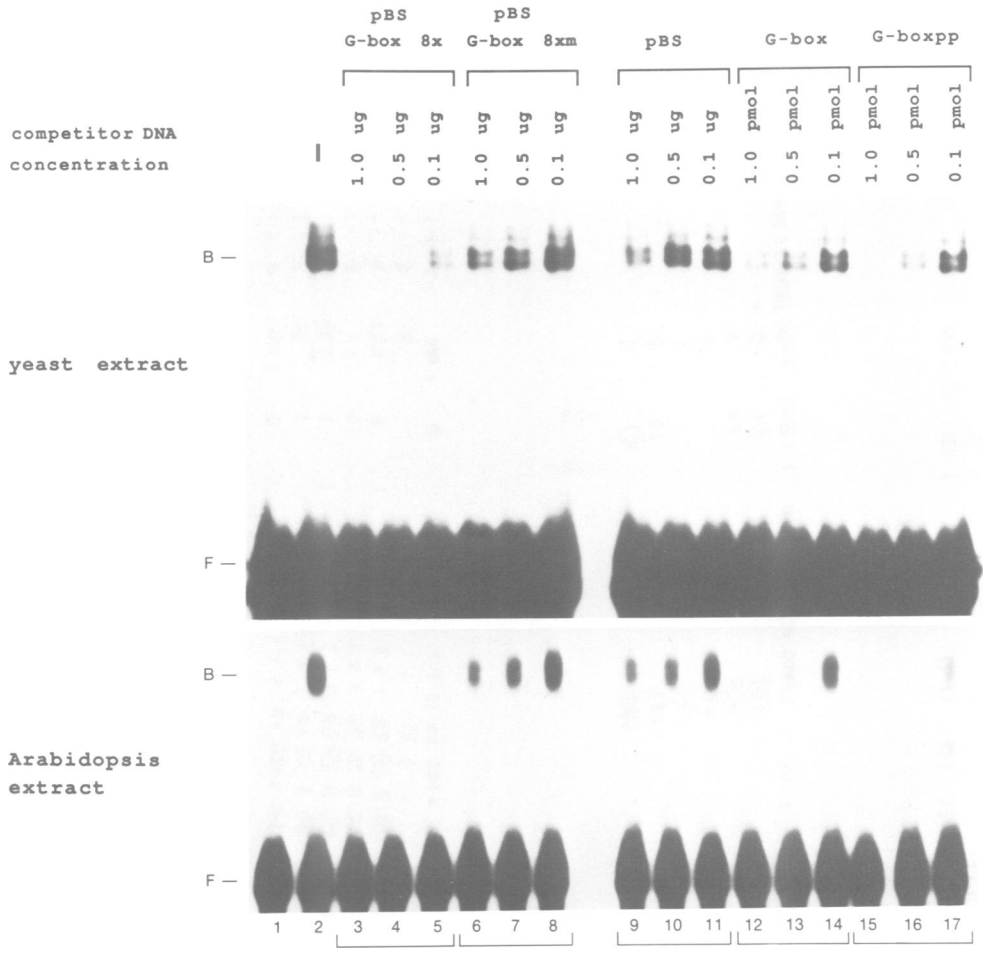

Fig. 5. Gel electrophoresis G box binding assay with *Arabidopsis* and yeast extract. The end-labeled *rbcS-1A* promoter fragment (−251 to −211) was incubated in the absence (**lane 1**) or presence (**lane 2**) of 4 μg of yeast (top panel) or plant (bottom panel) nuclear protein. Only one retarded complex is visible with *Arabidopsis* extract, whereas at least two closely spaced but distinct bands can be resolved with yeast extract. Competition for binding is observed with increasing amounts of specific competitor DNA containing the tomato *rbcS-3A* G box polymerized in pBS [**lanes 3–5**; 'pBS(G box)8X', 0.1, 0.5, 1.0 μg], as a 25 bp oligonucleotide (**lanes 12–14**; 'G box', 0.1, 0.5, 1.0 pmol) or as a modified 25 bp oligonucleotide in which a single base pair substitution creates a G box with a perfect palindrome (**lanes 15–17**; 'G boxpp', 0.1, 0.5, 1.0 pmol). In contrast, non-specific competitor DNA, containing the mutated polymerized tomato G box subcloned in pBS [**lanes 6–8**; 'pBS(G box)8XM'] or pBS plasmid alone (**lanes 9–11**), did not compete effectively. The DNA sequences of oligonucleotides used are shown in Figure 2A. The mobility of bound probe (B) or free probe (F) is indicated.

involved in the binding of plant and yeast GBF activities, methylation interference experiments were performed. A labeled *Arabidopsis rbcS-1A* fragment spanning −320 to −180 and containing the G box, both I boxes and a putative GT-1 box, was partially methylated, incubated with plant or yeast extracts and bound DNA fragments separated from free fragments by native polyacrylamide electrophoresis. After recovery of 'factor bound' and 'factor free' DNA fractions and cleavage with piperidine, products were analyzed on a sequencing gel (Figure 6). The methylation interference patterns generated by plant and yeast factors were indistinguishable. In both cases, DNA molecules carrying a methyl group at the indicated G residues were under-represented in the protein-complexed fraction compared with the unbound fraction. The G residues at these positions correspond precisely to the G box indicating that modification of this nucleotide inhibits the formation of a stable protein complex with both the plant and the yeast factors. No differences in methylation pattern between bound and unbound fractions were observed in the regions corresponding to either I box.

From these studies we conclude that a yeast nuclear protein (yGBF) specifically interacts with the G box present in plant promoters.

Discussion

Plant G box sequences are active in yeast and are bound in vitro by yGBF

In the first experiments described in this report we identified fragments of the *rbcS-1A* promoter which were capable of activating expression from a truncated CYC1 promoter in *Saccharomyces cerevisiae*. Mutation of two conserved

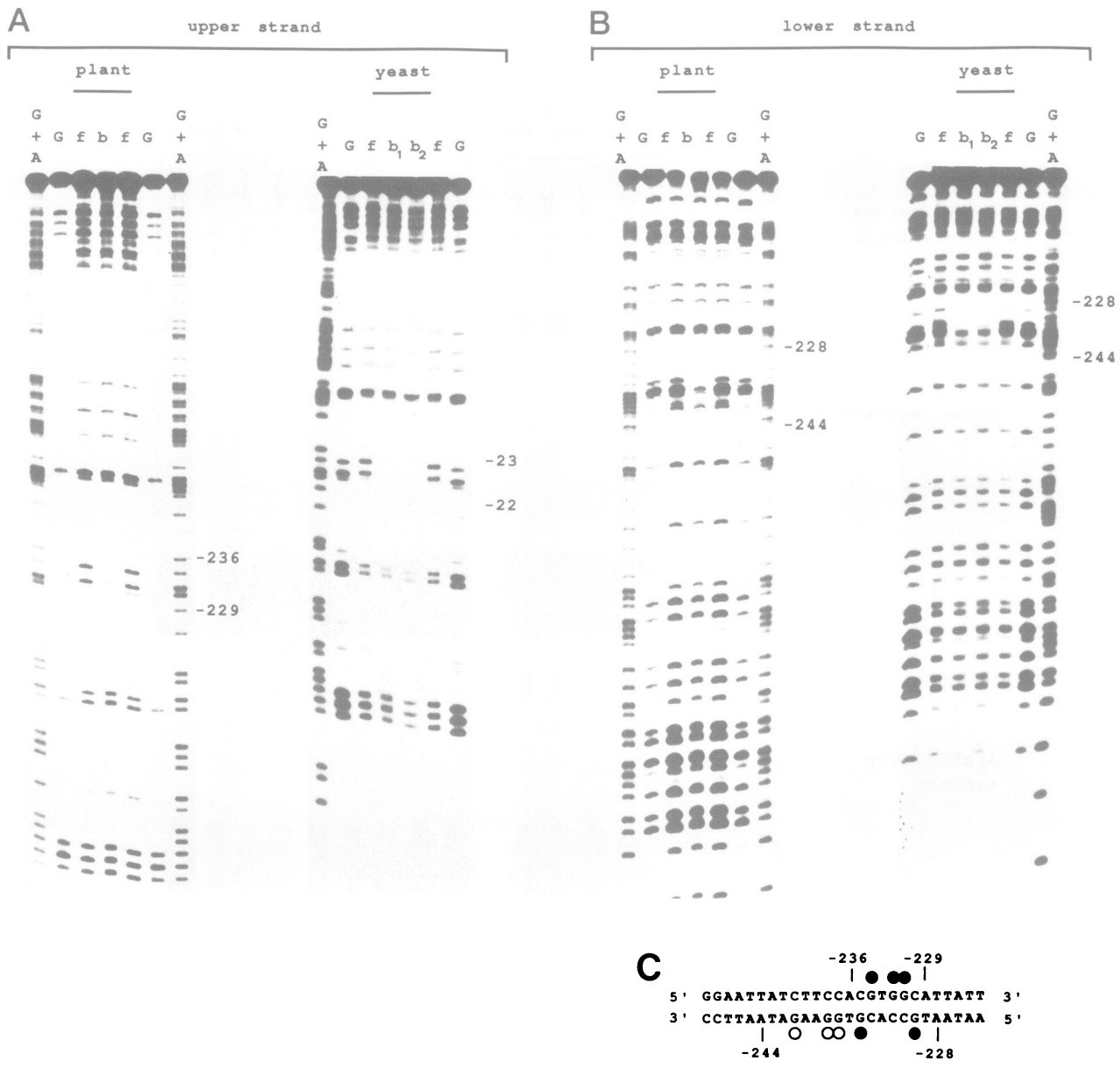

Fig. 6. Methylation interference experiment with *Arabidopsis* and yeast nuclear extract. Either the upper (coding) strand (**A**) or the lower (non-coding) strand (**B**) of the *Arabidopsis rbcS-1A* promoter fragment (−320 to −180) was labeled, incubated with plant or yeast extract and bound DNA bands eluted after non-denaturing gel electrophoresis. In addition to the G box, this fragment contains I box sequences recognized by GA-1 (both GATAAG motifs) and a putative GT-1 sequence (Donald and Cashmore, 1990). Eluted bands were cleaved with piperidine and the fragments resolved on an 8% polyacrylamide−8 M urea sequencing gel. In the scaled up binding reaction with yeast extract (see Materials and methods) only two of the retarded bands were visible (see Figure 5), both were eluted and they resulted in the same interference pattern (b1 and b2). In the case of the plant extract only one retarded band was observed. Lanes G and G+A correspond to G and G+A Maxam and Gilbert (1980) sequencing reactions, respectively. (**C**) Summary of the results obtained from the methylation interference analysis. Solid circles represent methylated G residues completely (open circles) or partially (closed circles) interfering with protein binding.

promoter elements found to be essential for *rbcS-1A* promoter function *in planta*, the G box and two I boxes, were also found to be required for expression mediated by a −390 to 60 *rbcS-1A* fragment in yeast. The requirement of both sequences for the expression mediated by this fragment indicates that expression is influenced by the interaction of the yeast transcriptional apparatus with both elements. In the case of the G box sequence, we have identified a yeast G box binding activity or activities (yGBF) which resembles plant nuclear factor GBF (Giuliano *et al.*, 1988) in *in vitro* binding specificity and which may be responsible for the observed G-box-dependent activation of the truncated CYC1 promoter *in vivo*. When excised from either the *Arabidopsis Adh*, *rbcS-1A* or tomato *rbcS-3A* promoters, the G box sequence was found to be a potent transcriptional activator of the truncated CYC1 promoter. We have also demonstrated I-box-dependent promoter activity in yeast. However, in this case the activity was considerably weaker than the G-box-dependent activity and we were unable to identify an associated yeast DNA binding activity.

Evidence for more than one yeast factor binding to G-box-like sequences

To identify previously described yeast promoter sequences that might potentially be bound by yGBF *in vivo*, we

Table I. G-box-like elements in 5′ non-coding regions of some yeast (*S.cerevisiae*) genes

Gene	Sequence[a]	Position (relative to ATG)	Reference
Acid phosphatase (PHO5)	CACACGTGGGA	−542	Arima *et al.* (1983)
AdhII	GTTCCACGTGA(rev.)	−783	Russell *et al.* (1983)
Alkaline phosphatase (PHO8)	GGGCCACGTGC(rev.)	−542	Kaneko *et al.* (1987)
Methionyl-tRNA-synthetase (cytoplasmic)	TCACGAGGCAC(rev.)	−189	Walter *et al.* (1983)
Methyl-transferase (PEM2)	CCACGTGGAAC	−484	Kodaki and Yamashita (1987)
Phosphoglucoisomerase	GTGCCACGTGAC(rev.)	−475	Tekamp-Olson *et al.* (1988)
Pyruvate kinase	CTGCCACGTGGG(rev.)	−527	Burke *et al.* (1983)
Ubiquinol-cytc-oxidoreductase	GTTCCACGTGA(rev.)	−245	Maarse and Grivell (1987)

[a]The Genetics Computer Group Sequence Analysis Software Package (GCG version 6.0) program 'Find' (Devereux *et al.*, 1984) was run on the Genebank database (release 59.0) to identify homology to the G box consensus sequence: 5′-CACGTGGCA, allowing for one mismatch. All sequences are written 5′ to 3′. Those labeled 'rev.' are from the bottom (non-coding) strand of the respective promoter sequence.

searched the Genebank database for sequences homolgous to the plant G box consensus C/A-CACGTGGCA. G box homology was found in 5′ non-coding regions of several yeast genes (Table I): acid phosphatase (Arima *et al.*, 1983); alkalkine phosphatase (Kaneko *et al.*, 1987); alcohol dehydrogenase II (Russell *et al.*, 1983); methyl-transferase (Kodaki and Yamashita, 1987); methionyl-tRNA synthetase (Walter *et al.*, 1983); phosphoglucoisomerase (Tekamp-Olson *et al.*, 1988); pyruvate kinase (Burke *et al.*, 1983); ubiquinol-cytc-oxidoreductase (Maarse *et al.*, 1987).

An interaction between one of the yeast promoter sequences, the acid phosphatase (PHO5) G-box-like promoter sequence, and a yeast transcription factor has been documented. The positively acting PHO4 regulatory protein, together with the PHO2 regulatory protein, mediates the derepression of the PHO5 gene under low phosphate conditions (Toh-e *et al.*, 1975; Bostian *et al.*, 1983; Sengstag and Hinnen, 1988; Vogel *et al.*, 1989). The PHO4 factor binds specifically *in vitro* to two regions of the PHO5 promoter, which when deleted specifically reduce PHO5 expression *in vivo* (Vogel *et al.*, 1989). One of these two PHO5 promoter regions shows homology with the plant G box (Table I). To determine if the PHO4 factor contributes to our observed G-box-dependent CYC1−*lacZ* expression in yeast, we investigated the expression of several CYC1−*lacZ* constructs in a PHO4 mutant. Relative to the intact CYC1 promoter and truncated CYC1 promoter controls, the CYC1−*lacZ* expression mediated by plant G box fragments was not significantly different in the PHO4 mutant and insensitive to the concentration of phosphate in the growth media (data not presented).

Other workers have also identified yeast DNA binding activities that resemble yGBF (Bram and Kornberg, 1987; Beckmann and Kadesch, 1989; Chodosh *et al.*, 1989; Vogel *et al.*, 1989), although the sites recognized by these activities do not match the G box as closely as the sequences shown in Table I. The yeast factor described by Chodosh *et al.* (1989) and Bram and Kornberg (1987) interacts with a near-palindromic sequence CCACGTGA, the binding site of the mammalian transcription factor USF or MLTF (Carthew *et al.*, 1985; Sawadogo and Roeder, 1985; Chodosh *et al.*, 1986). Beckmann and Kadesch (1989) identified a yeast factor with similar properties to NFμE3, a mammalian DNA binding protein interacting with the E3 site of the immunoglobulin heavy chain enhancer. The E3 site (TCATGTGG) differs from the G box by two C−T transitions. Like the plant G box sequence, this motif can also activate expression in yeast from the truncated CYC1−*lacZ* construct, although its activity is significantly lower than the G box sequences described here (Beckmann and Kadesch, 1989).

In this report we have presented the first evidence for the existence of a yeast homologue to a plant DNA binding protein. The similarities in DNA sequence recognition by plant and yeast proteins suggest an evolutionary relatedness at the level of the DNA binding domain. We do not know at this time the extent to which the apparent similarities of the plant and yeast GBF activities we have described can also be extended to mammalian factors which bind to G-box-like motifs. More definitive comparisons will be possible with a more detailed investigation of the binding site requirements of purified DNA binding activities.

Materials and methods

Yeast strains, media and transformation

Yeast strains used in this study were 146a (**a** *ura3 his4 leu2 trp1*), 29α (α *ura3 his4 leu2 trp1*), YPH49 a/α (**a**/α, *ura3/" lys2/" ade2/" trp1/"*), RK4-TU5 (**a** *his3 trp1 ura3 pho4 pho3*) and DBY747 (**a** *his3 trp1 ura3 leu2*). Cells were grown at 30°C in a minimal medium of yeast nitrogen base without amino acids (Difco) supplemented with 2% glucose, and 40 μg/ml of the required amino acids. For phosphate regulation experiments, low-P_i and high-P_i minimal media of Meyhack *et al.* (1982) was used. Yeast was transformed by the lithium acetate procedure (Ito *et al.*, 1983) with selection of transformants on minimal media lacking exogenous uracil.

Plasmid construction

The 2μ-based yeast vectors pLG669-Z and pLG670-Z containing the CYC1−*lacZ* and the UAS-deleted CYC1−*lacZ* fusions, respectively, have been described (Figure 1A; Guarente and Ptashne, 1981). pLG670-Z, which is derived from pLG669-Z by removal of a 450 bp CYC1 *Xho*I fragment containing UAS elements, was used to subclone all the heterologous promoter fragments described. Digestion of pLG670-Z with *Sal*I and *Xho*I removes remaining upstream CYC1 sequences and facilitates the subcloning of heterologous fragments at a position ~180 bp from the most proximal CYC1 transcriptional start site.

Oligonucleotides containing G box and I box binding-site motifs were subcloned into pBSBgl, a modified version of the plasmid pBS-SK(+) (Stratagene) in which a *Bgl*II linker was inserted into the *Eco*RV site, and polymerized by successive cycles of *Sca*I−*Bam*HI and *Sca*I−*Bgl*II digestion and ligation. Single or multiple directly repeated copies were removed with *Sal*I and *Bam*HI for yeast expression vector subcloning. As a consequence of this subcloning, the polymerized sequences were flanked by 26 and 22 nucleotides of pBSBgl polylinker on 3′ and 5′ sides, respectively.

Arabidopsis Adh and *rbcS-1A* G box sequences were excised from replicative-form M13mp18 subclones modified by site-directed mutagenesis (Kunkel, 1985; Donald and Cashmore, 1990): 10 bp substitution mutations flanking the *Adh* G box (introducing *Xho*I/*Bgl*II sites) and 12 bp substitutions

flanking the *rbcS-1A* G box (introducing *Eco*RV/*Bgl*II sites). A −251 to −211 *rbcS-1A* fragment (Figure 1B, fragments 11 and 12) and a −220 to −187 *Adh* fragment (fragment 14) containing G box sequences were excised with *Bgl*II and subcloned into the *Bam*HI site of pUC19 and *Eco*RV site of pBS-SK(+) to form plasmids pRG40 and pAG33, respectively. For subsequent yeast expression vector subcloning, these plasmids were digested with *Sma*I and *Sal*I and insert fragments ligated with pLG670-Z linearized with *Sal*I (Klenow filled) and *Xho*I. In the resulting yeast expression vector subclones genomic sequences were flanked by 4 and 8 bp of pUC19 polylinker on 3′ and 5′ sides, respectively, in the case of *rbcS-1A* G box construct and 18 bp of pBS-SK(+) polylinker on either side in the case of the *Adh* G box construct.

Larger *Arabidopsis Adh* and *rbcS-1A* promoter fragments were subcloned into *Sal*I/*Xho*I-treated pLG670-Z from plasmids described by Donald and Cashmore (1990). *Adh* fragment 13 (Figure 1C) was subcloned from plasmid pAdh11 by use of the genomic −125 bp *Sal*I site and a 5′ *Bam*HI polylinker site. *RbcS-1A* fragments 3, 4, 5, 7 and 8 were cloned into pLG670-Z after initial pBS-SK(+) subcloning from replicative form M13mp18 containing the −1700 to +21 *rbcS-1A* promoter or mutated derivatives. In these cases, the −60 bp *rbcS-1A* genomic *Hin*dIII site was fused directly to the −178 bp *Xho*I site of pLG670-Z. Fragment 6 was separated from the truncated CYC1 promoter by 6 bp of pUC19 polylinker as a consequence of subcloning from plasmid pRΔ5032; fragments 9 and 10 which span the *rbcS-1A* LRE were separated by 14 bp of non-genomic plasmid polylinker as a result of subcloning from plasmids pIGL1 and pIGL2.

β-Galactosidase assays

Crude yeast extracts were prepared from exponentially growing yeast transformants and β-galactosidase activity normalized to the amount of protein assayed as described (Bradford, 1976; Rose and Botstein, 1983). Units were calculated according to Miller (1972). At least three independent transformants of each construct were included in the analysis.

RNase mapping

Total RNA was isolated from 20 ml yeast cell cultures grown in minimal selective media (−uracil) as described (Guarente and Mason, 1983) except that RNA was purified further by precipitation with lithium chloride. RNA complementary to CYC1−*lacZ* was synthesized *in vitro* by SP6 polymerase (Promega) using [α-^{35}S]UTP (1000 Ci/mmol, Amersham) and a *Bam*HI-linearized pGEM4 plasmid template containing a 210 bp *Sph*I−*Pvu*II DNA fragment excised from plasmid pLG669-Z which spans the CYC1−*lacZ* fusion junction. RNase A/T1 mapping was performed as described (Ausubel *et al.*, 1987). Between 5 and 10 μg of yeast RNA was used for each assay and protected fragments analyzed on 6% polyacrylamide−8 M urea gels.

Preparation of nuclear extract

Plant nuclear extracts were prepared from 4-week-old *Arabidopsis* seedlings (Columbia) grown under greenhouse conditions, as described by Giuliano *et al.* (1988). Yeast nuclear proteins were isolated from the strain YPH146a by the method of Wiederrecht *et al.* (1987) with modifications described by Beckmann and Kadesch (1989).

Gel retardation assay

The −251 to −211 *rbcS-1A* G box fragment was used for gel retardation assays. The fragment was prepared by *Hin*dIII−*Eco*RI digestion of pRG40 (see above) and was end-labeled with [α-^{32}P]dATP and DNA polymerase (Klenow). Binding reactions were carried out for 30 min at room temperature in a total volume of 20 μl. A typical reaction contained 10 mM Tris−HCl, pH 7.5, 40 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, 4% glycerol, 5 μg poly(dI−dC), 10 000 c.p.m. radioactively labeled DNA, 2 μg nuclear extract and specific competitor DNA as indicated, respectively. After incubation, the reactions were loaded on a 5% polyacrylamide gel (38:2) in 25 mM Tris−HCl, 190 mM glycine, 1 mM EDTA, pH 8.3, which was pre-run for 1−2 h at 6 V/cm. Electrophoresis was carried out at the same voltage. Gels were dried and exposed overnight to X-ray films.

Methylation interference experiment

The −320 to −180 *rbcS-1A* DNA fragment used in the methylation interference experiment contains both I boxes, the G box and a putative GT-1 box with homology to pea *rbcS-3A* box III* (Donald and Cashmore, 1990). The DNA fragment, labeled on either end after *Hin*dIII or *Eco*RI digestion (5×10^5 c.p.m.), was partially methylated with dimethyl sulfate in the presence of 1 μg pUC DNA according to Maxam and Gilbert (1980). After ethanol precipitation, the binding reactions were carried out as described above, except that the reaction volume was scaled up 4-fold, the amount of poly(dI−dC) and nuclear extract 10-fold. The free and bound DNA fragments were separated on a non-denaturing 5% polyacrylamide gel. After autoradiography of the wet gel overnight, the prominent shifted and 'free' bands were excised and recovered by electroelution on DEAE paper, eluted for 2 h at 68°C in electroelution buffer (20 mM Tris−HCl, pH 8.0, 0.1 mM EDTA, 1.5 M NaCl), phenol chloroform (1:1) extracted, ethanol precipitated and treated with 100 μl 10% piperidine for 30 min at 90°C. The samples were lyophilized overnight and equal amounts of radioactivity were subjected to electrophoresis through a 8% polyacrylamide−8 M urea sequencing gel.

Acknowledgements

We thank Drs H.Beckmann and T.Kadesch for useful discussions, for informing us of results prior to publication and for their critical review of the manuscript for which we also thank Drs M.Ahmad and A.Sagar. Dr Beckmann provided plasmids pLG669-Z and pLG670-Z used in the yeast expression experiments and the SP6 plasmid subclone used to generate anti-sense RNA in RNase protection experiments. We also thank Drs K.Tatchell and S.Parent for yeast strains. For help with the DNA sequence database search we thank Dr Peter Bringmann. This work was supported by an NIH grant GM-38409 (to A.R.C.), a DOE grant DE-FG02-87ER13680 (to A.R.C.) and a post-doctoral NSF fellowship (to R.G.K.D.)

References

Arima,K., Oshima,T., Kubota,I., Nakamura,N., Mizunaga,T. and Toh-e,A. (1983) *Nucleic Acids Res.*, **11**, 1657−1672.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (eds) (1987) *Current Protocols in Molecular Biology*. Wiley, New York.

Beckmann,H. and Kadesch,T. (1989) *Mol. Cell. Biol.*, **9**, 4535−4540.

Bostian,K.A., Lemire,J.M. and Halvorson,H.O. (1983) *Mol. Cell. Biol.*, **3**, 839−853.

Bram,R.J. and Kornberg,R.D. (1987) *Mol. Cell. Biol.*, **7**, 403−409.

Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248−254.

Buratowski,S., Hahn,S., Sharp,P.A. and Guarente,L. (1988) *Nature*, **334**, 37−43.

Burke,R.L., Tekamp-Olson,P. and Najarian,R. (1983) *J. Biol. Chem.*, **258**, 2193−2201.

Carthew,R.W., Chodosh,L.A. and Sharp,P.A. (1985) *Cell*, **43**, 439−448.

Castresana,C., Staneloni,R.E., Malik,V.S. and Cashmore,A.R. (1987) *Plant Mol. Biol.*, **10**, 117−126.

Cavallini,B., Huet,J., Plassat,J.-L., Sentenac,A., Egly,J.-M. and Chambon,P. (1988) *Nature*, **334**, 77−80.

Chang,C. and Meyerowitz,E.M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1408−1412.

Chodosh,L.A., Carthew,R.W. and Sharp,P.A. (1986) *Mol. Cell. Biol.*, **6**, 4723−4733.

Chodosh,L.A., Oleson,J., Hahn,S., Baldwin,A.S., Guarente,L. and Sharp,P.A. (1988) *Cell*, **53**, 11−24.

Chodosh,L.A., Buratowski,S. and Sharp,P.A. (1989) *Mol. Cell. Biol.*, **9**, 820−822.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387−395.

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

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

Fischer,J.A., Gininger,E., Maniatis,T. and Ptashne,M. (1988) *Nature*, **332**, 853−856.

Gidoni,D., Brosio,P., Bond-Nutter,D., Bedbrook,J. and Dunsmuir,P. (1989) *Mol. Gen. Genet.*, **215**, 337−344.

Green,P.J., Kay,S.A. and Chua,N.-H. (1987) *EMBO J.*, **6**, 2543−2540.

Green,P.J., Yong,M.-H., Cuozzo,M.Y., Murakami,K., Silverstein,P. and Chua,N.-H. (1989) *EMBO J.*, **7**, 4035−4044.

Guarente,L. (1983) *Methods Enzymol.*, **101**, 181−191.

Guarente,L. and Ptashne,M. (1981) *Proc. Natl. Acad. Sci. USA*, **87**, 2199−2203.

Guarente,L. and Mason,T. (1983) *Cell*, **32**, 1279−1286.

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

Harshman,K.D., Moye-Rowley,W.S. and Parker,C.S. (1988) *Cell*, **53**, 321−330.

Hayes,T.E., Sengupta,P. and Cochran,B.H. (1988) *Genes Dev.*, **2**, 1713−1722.

Horikoshi,M., Carey,M.F., Kakidani,H. and Roeder,R.G. (1988) *Cell*, **54**, 665−669.

Ito,H., Fukuda,K.Y., Murata,K. and Kimura,A. (1983) *J. Bacteriol.*, **153**, 163−168.

Jones,R.H., Moreno,S., Nurse,P. and Jones,N.C. (1988) *Cell*, **53**, 659–667.
Kakidani,H. and Ptashne,M. (1988) *Cell*, **52**, 161–167.
Kaneko,Y., Hayashi,N., Toh-e,A., Banno,I. and Oshima,Y. (1987) *Gene*, **58**, 137–148.
Kodaki,T. and Yamashita,S. (1987) *J. Biol. Chem.*, **262**, 15428–15435.
Kuhlemeier,C., Fluhr,R., Green,P.J. and Chua,N.-H. (1987) *Genes Dev.*, **1**, 247–255.
Kuhlemeier,C., Cuozzo,M., Green,P.J., Goyvaerts,E., Ward,K. and Chua,N.-H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4662–4666.
Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.
Lam,E., Benfey,P.N.,Gilmartin,P.M., Fang,R.-X. and Chua,N.-H. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7890–7894.
Lin,Y.-S. and Green,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 109–113.
Lin,Y.-S., Carey,M.F., Patashne,M. and Green,M. (1988) *Cell*, **54**, 659–664.
Ma,J., Przibilla,E., Hu,J., Bogorad,L. and Ptashne,M. (1988) *Nature*, **334**, 631–633.
Maarse,A.C. and Grivell,L.A. (1987) *Eur. J. Biochem.*, **165**, 419–425.
Manzara,T. and Gruissem,W. (1988) *Photosynth. Res.*, **16**, 117–130.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Meyhack,B., Bajwa,W., Rudolph,H. and Hinnen,A. (1982) *EMBO J.*, **1**, 675–680.
Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Rose,M. and Botstein,D. (1983) *Methods Enzymol.*, **101**, 167–180.
Russell,D.W., Smith,M., Williamson,V.M. and Young,E.T. (1983) *J. Biol. Chem.*, **258**, 2674–2682.
Sawadogo,M. and Roeder,R.G. (1985) *Cell*, **43**, 165–175.
Schulze-Lefert,P., Dangl,J.L., Becker-André,M., Hahlbrock,K. and Schulz,W. (1989) *EMBO J.*, **8**, 651–656.
Senstag,C. and Hinnen,A. (1988) *Gene*, **67**, 223–228.
Staiger,D., Kaulen,H. and Schell,J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6930–6934.
Struhl,K. (1987) *Cell*, **50**, 841–846.
Tekamp-Olson,P., Najarian,R. and Burke,R.L. (1988) *Gene*, **73**, 153–161.
Toh-e,A., Kakimoto,S. and Oshima,Y. (1973) *J. Bacteriol.*, **113**, 727–738.
Vogel,K., Hoerz,W. and Hinnen,A. (1989) *Mol. Cell. Biol.*, **9**, 2050–2057.
Walter,P., Gangloff,J., Bonnet,J., Boulanger,Y., Ebel,J.-P. and Fasiolo,F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2437–2441.
Webster,N., Jin,J.R., Green,S., Hollis,M. and Chambon,P. (1988) *Cell*, **52**, 169–178.
Wiederrecht,G., Shuey,D.J., Kibbe,W.A. and Parker,C.S. (1987) *Cell*, **48**, 507–515.

Received on January 22, 1990; revised on March 5, 1990