Mol Gen Genet (1999) 261: 635–643

ORIGINAL PAPER

P. B. F. Ouwerkerk · J. Memelink

A G-box element from the *Catharanthus roseus* strictosidine synthase (*Str*) gene promoter confers seed-specific expression in transgenic tobacco plants

Received: 6 August 1998 / Accepted: 22 January 1999

Abstract The enzyme encoded by the strictosidine synthase (*Str*) gene from *Catharanthus roseus* catalyses a key step in the biosynthesis of the pharmaceutically important terpenoid indole alkaloids. *Str* cDNA and genomic clones have already been isolated, allowing us to study the regulation of *Str* gene expression. Here we focus on the role of a putative *cis*-acting element, CACGTG, in the *Str* promoter. This sequence is known as a G-box, and functions as a transcription-regulating sequence in a number of other promoters. By means of electrophoretic mobility shift assays it was demonstrated that the *Str* G-box is capable of interacting with nuclear factors in tobacco and with the cloned tobacco G-box-binding factor TAF-1. Disruption of the *Str* G-box sequence by two single-nucleotide mutations prevented binding of factors, thereby demonstrating the specificity of the observed interactions. Functional analysis in transgenic tobacco plants demonstrated that these mutations also reduced the transcriptional activity of constructs containing tetramers of the *Str* G-box sequence. Expression directed by a tetramer of the *Str* G-box fused to a truncated promoter containing only a TATA box was confined to seeds and was found to increase during seed maturation. Thus, the *Str* G-box tetramer is able to direct seed-specific expression independently of other regulatory sequences. G-box-directed expression in leaves required the presence of an enhancer region from the cauliflower mosaic virus (CaMV) 35S promoter. The results indicate that the G-box needs to interact with other elements to drive expression in leaf, and that it can by itself confer seed-specific expression as a multimer. The fact that only some of the G-boxes found in different promoters serve as seed-specific elements indicates that sequences flanking the G-box determine the transcriptional activity in different tissues. Based on sequence comparisons we propose that the nucleotides at positions −4, −3, −2 and/or +4 are important in determining seed-specific expression.

Key words *Catharanthus roseus* · G-box · Seed-specific expression · *Str* gene · TAF-1

Communicated by A. Kondorosi

P. B. F. Ouwerkerk · J. Memelink (✉)
Institute of Molecular Plant Sciences
Leiden University, Clusius Laboratory
PO Box 9505, 2300 RA Leiden, The Netherlands
e-mail: memelink@rulbim.leidenuniv.nl
Tel.: +31-71-5274751; Fax: +31-71-5275088

Introduction

An important aspect of the regulation of gene expression is the specific interaction of transcription factors with *cis*-acting elements present in gene promoter regions. These *cis*-acting elements determine gene activity during plant development, and in response to external stimuli such as hormones, pathogens or light. Functional studies on plant promoters have resulted in the identification of a number of *cis*-acting elements and their cognate *trans*-acting DNA-binding factors (reviewed by Foster et al. 1994; Menkens et al. 1995; Meshi and Iwabuchi 1995).

We are interested in the transcriptional regulation of genes involved in the biosynthesis of the pharmaceutically important terpenoid indole alkaloids in the tropical plant *Catharanthus roseus* (Madagascar periwinkle, family Apocynaceae). This group of secondary metabolites is thought to have anti-herbivore activity and to function in protection against pathogens and/or UV radiation (Aerts et al. 1991; Hartmann 1991; Luijendijk 1995; Thomas et al. 1995). One of the key steps in the biosynthesis of terpenoid indole alkaloids is the condensation of the amino acid derivative tryptamine and the terpenoid secologanin, catalysed by the enzyme strictosidine synthase (EC 4.3.3.2.) (reviewed in Meijer et al. 1993; Hashimoto and Yamada 1994). The product of this reaction is strictosidine, which is the precursor for all other terpenoid indole alkaloids. cDNA and genomic clones of *Str* have been obtained from *C. roseus*, al-

lowing studies of the molecular mechanisms that control *Str* gene expression (McKnight et al. 1990; Pasquali et al. 1992, 1999). In *C. roseus* the *Str* gene was shown to be expressed in various tissues, with highest levels in roots. *Str* was also found to be induced by fungal elicitors, consistent with its presumed role in plant defence.

A search of the *Str* promoter sequence for putative *cis*-regulatory DNA-sequences that might be involved in determining the *Str* expression pattern revealed the presence of the palindromic sequence CACGTG, which is known as the G-box. The G-box was originally found in *RbcS* promoters from various plants species, as a conserved region which is protected by plant nuclear extracts from DNase digestion in vitro (Giuliano et al. 1988). Since then, G-box and G-box-like sequences have been studied for their functional significance in vivo; they have been shown to be involved in the regulation of a variety of unrelated genes in various responses, following exposure to visible and UV light (Schulze-Lefert et al. 1989a, 1989b; Block et al. 1990; Chattopadhyay et al. 1998), dehydration-stress (Lam and Chua 1991; Dolferus et al. 1994), cold-stress (Dolferus et al. 1994), abscisic acid (Marcotte et al. 1989; Lam and Chua 1991) and sucrose (Urwin and Jenkins 1997). Moreover, G-box(-like) sequences were also found to determine tissue-specific expression patterns, for example in seeds and roots (Salinas et al. 1992; Thomas 1993).

A number of cDNAs encoding nuclear proteins which specifically interact with G-box(-like) sequences have been identified. Nearly all of these proteins are members of the basic leucine zipper (bZIP) family (Guiltinan et al. 1990; Oeda et al. 1991; Weisshaar et al. 1991; Schindler et al. 1992; Menkens and Cashmore 1994; Meier and Gruissem 1994; Hong et al. 1995; Izawa et al. 1994; for a review see Menkens et al. 1995). In addition, a *trans*-acting factor of the MYC type has been shown to interact with a G-box sequence (De Pater et al. 1997).

It is still not clear how identical G-box sequences can be involved in the specific responses of a wide variety of unrelated genes to a number of completely different stimuli. One important factor to consider is that the affinities of G-box binding *trans*-acting factors (GBFs) are also dependent on the sequence context flanking the G-box (Williams et al. 1992; Izawa et al. 1993). This is demonstrated by the fact, among others, that some perfect G-boxes have been found not to bind GBFs at all, whereas other G-boxes form either characteristic low- (type A) or high (type B)-mobility complexes in electrophoretic mobility shift assays (Williams et al. 1992). Several promoter studies that have focused on the function of G-box sequences in vivo have revealed that additional *cis*-acting elements are necessary for *trans*-activation of gene expression, thereby demonstrating the existence of synergism between certain G-boxes and other *cis*-acting elements (Donald and Cashmore 1990; Loake et al. 1992; Kawagoe et al. 1994; Shen and Ho 1995). Based on the data currently available, the sequence context of a particular G-box is insufficient to allow one to predict whether that G-box is recognized by GBFs, whether it is a functional *cis*-acting element, and which in vivo activity it may confer.

The aim of the experiments described here was to determine whether the G-box in the *C. roseus Str* promoter binds GBFs and whether this element can drive gene expression. Previous studies showed that 0.6 kb of *Str* upstream sequences could confer expression on the *gusA* reporter gene in transgenic tobacco plants (Pasquali et al. 1994, 1999). Here we show that the *Str* G-box is recognized in vitro by nuclear factors from tobacco. Furthermore, we demonstrate that a tetramer of the *Str* G-box directs seed-specific expression when combined with a truncated promoter containing only a TATA box, and elevates the expression level in leaves in a synergistic interaction with the *as-1* enhancer-containing region of the CaMV 35S promoter.

Materials and methods

Vector construction and generation of transgenic plants

Head-to-tail tetramers of the wild-type (4SW) and mutant (4SM) *Str* G-box were made using 18-bp oligonucleotides flanked by *Sal*I and *Xho*I sites (underlined in Fig. 1A), as described by Ouwerkerk and Memelink (1997), and were subsequently cloned upstream of the truncated CaMV 35S −47 and −90 promoters using the *Sal*I sites in the plasmids GusSH-47 and GusSH-90 (Pasquali et al. 1994). These vectors contain the −47 to +27 and −90 to +27 sequences, respectively, from the CaMV 35S promoter, in a

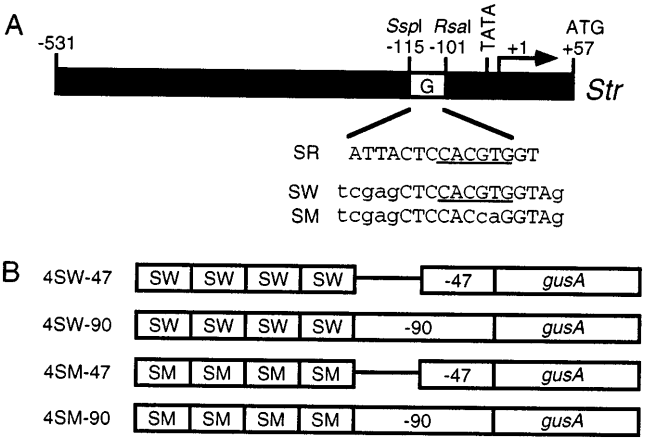

Fig. 1 **A** Schematic representation of the *Str* –531 promoter from *C. roseus*. Indicated are the TATA box (–29 to –26), the mRNA transcription start site at +1 and the translation start codon at +57. The *Ssp*I and *Rsa*I restriction sites generate the G-box-containing fragment SR (–115 to –101) which was used in EMSAs. The sequence of SR is presented and the G-box is *underlined*. The nucleotide sequence of the *Str* gene is available from the EMBL/GenBank/DDBJ database under accession number Y10182. The synthetic oligonucleotides SW and SM represent wild-type and mutant *Str* G-box sequences, respectively. The sequences in *lower case* at the 5′ and 3′ ends represent *Sal*I and *Xho*I sites used for tetramerization. The GT to CA mutation in the G-box is also shown in *lower case*. **B** Structure of the chimeric *Str* G-box-*gusA* fusion constructs. Tetramers of wild type (SW) or mutant (SM) *Str* G-box sequences were cloned upstream of the truncated CaMV 35S –90 or –47 promoters

transcriptional fusion to the *gusA* reporter gene. Next, the tetramer-*gusA* fusion constructs were cloned as *Xba*I-*Xho*I fragments in the corresponding restriction sites of the binary vector pMOG$\lambda\varpi$ CAT (Pasquali et al. 1994), resulting in the constructs 4SW-47, 4SM-47, 4SW-90 and 4SM-90. The *gusA* constructs were transferred via *Agrobacterium tumefaciens* LBA4404 to tobacco (cv. Petit Havana SR1), using the leaf disc transformation procedure (Horsch et al. 1991). Transgenic plants were selected for growth on 100 µg/ml of each of the antibiotics kanamycin, vancomycin and cefotaxime, on MS medium supplemented with 3% sucrose and 0.7% Plant Tissue Culture agar. Greenhouse-grown plants were used for seed production. Growth conditions in the greenhouse and in the growth chamber were 21° C and a photoperiod of 16 h light (2000 lx) per day.

Electrophoretic mobility shift assays (EMSAs)

Nuclear extracts were prepared from mature tobacco leaves as described by Green et al. (1991). Crude *E. coli* extract containing TAF-1, which was expressed from the corresponding cDNA clone, was prepared as described by Oeda et al. (1991). All EMSA reactions contained 3 µg of poly(dIdC)-poly(dIdC) (Pharmacia LKB Biotechnology) and 0.025 ng (SR) or 0.5 ng (4SW) of ^{32}P end-labelled probe (10^8 cpm/µg) in nuclear extraction buffer, in a total volume of 10 µl (Green et al. 1991). The amounts of competitor DNAs and protein extracts used are indicated in the text. The structures and sequences of competitor fragments and probes used are presented in Fig. 1A. The reactions were incubated for 20 min at room temperature and subsequently loaded, with the power switched on (10 V/cm), onto 5% acrylamide/bisacrylamide (37.5:1) gels in 0.5 × TBE as running buffer. The gels were dried on DE81 paper (Whatman) and autoradiographed.

Preparation of leaf and seed extracts and β-glucuronidase (GUS) assays

Protein extracts were prepared by homogenization in GUS extraction buffer of mature seeds or leaf discs from 6- to 8-week-old transgenic tobacco plants, using a Potter S homogenizer (Braun). GUS enzyme activities were measured according to Jefferson (1987) using a Perkin Elmer LS50B fluorimeter, and protein concentrations were determined by the Bradford (1976) method.

Histochemical detection of GUS activity

Histochemical detection of GUS activity with X-Gluc as substrate was essentially carried out as described previously by Jefferson (1987). Flowers of mature greenhouse-grown primary transformant tobacco plants were labelled when the petals were fully expanded (0 DAF). Seed capsules were harvested at 5 DAF (days after flowering), 10 DAF and 15 DAF and cross sections of about 3 mm thickness were prepared. The seeds were immobilized in the capsules by pipetting several drops of 0.8% agarose (in 100 mM sodium phosphate buffer pH 7.0, 10 mM EDTA) onto the surface of the cross sections. After incubation with X-Gluc, the cross sections were treated with 4% hypochlorite to reduce browning of the tissues.

Seedlings were grown on MS medium supplemented with 3% sucrose and 0.7% Plant Tissue Culture agar, and directly incubated with X-Gluc without further treatment. Chlorophyll was removed by washing with 70% ethanol.

RNA extraction and Northern analysis

RNA was isolated from various *C. roseus* tissues using the hot phenol-LiCl procedure (Van Slogteren et al. 1983) – except for RNA from seeds, which was extracted using the guanidinium protocol described by Chomczynski and Sacchi (1987). Northern hybridization analysis was performed as described by Memelink et al. (1994). Integrity and equal loading of RNA samples was verified by ethidium staining of ribosomal RNA bands.

Results

Binding of tobacco nuclear proteins and recombinant tobacco TAF-1 to the *Str* G-box

The palindromic G-box sequence CACGTG in the *Str* promoter is located between positions −108 and −103 relative to the mRNA transcription start site. To investigate the functionality of this sequence element, both in vitro and in vivo approaches were adopted. EMSAs were used to characterize in vitro binding of tobacco nuclear proteins to the *Str* G-box, which was used in both monomeric (SR) and tetrameric (4SW) forms. A mutant tetramer (4SM) contained two point mutations which disrupt the palindromic sequence of the G-box element (Fig. 1A). As shown in Fig. 2A, incubation of 20 µg of tobacco nuclear extract with radioactively labelled 4SW probe resulted in formation of a retarded complex. The migration behaviour of the complex resembled that of the A-type G-box binding activity described by Williams et al. (1992). The complex could be competed to a large degree with unlabelled 4SW, whereas the same amount of the mutant 4SM tetramer gave no competition.

EMSAs carried out with extracts prepared from *E. coli* cells in which the tobacco G-box binding factor

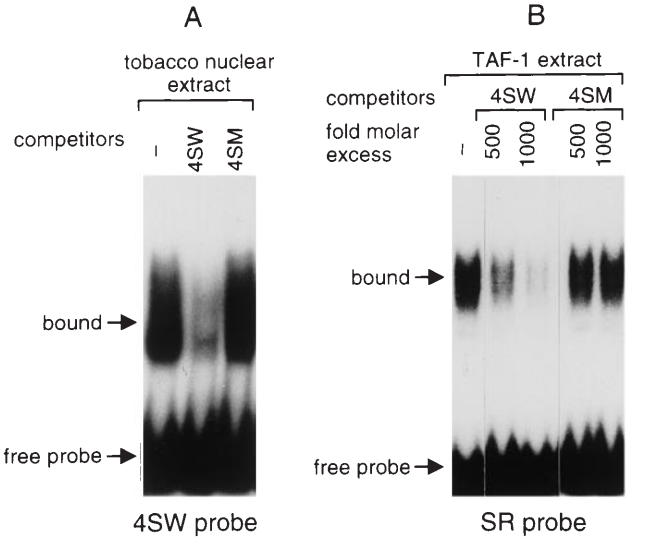

Fig. 2 **A** Binding of tobacco nuclear proteins to a tetramer (4SW) of the *Str* G-box. Binding reactions contained 20 µg of tobacco nuclear extract and 0.5 ng of the 4SW probe. Competitor fragments were included in the molar excess as indicated. **B** Binding of tobacco nuclear factor TAF-1 to a monomer of the *Str* G-box (fragment SR). Binding reactions contained 0.025 ng of probe and 0.5 µg of an *E. coli* protein extract containing TAF-1, expressed from its corresponding cDNA clone (Oeda et al. 1991). Competitor fragments were included in excess as indicated

TAF-1 was expressed from its corresponding cDNA clone (Oeda et al. 1991) gave results similar to those obtained with tobacco nuclear extract. Incubation of the monomer probe SR with extract containing TAF-1 gave rise to a TAF-1/G-box complex which could be competed with 4SW, but not with the mutant sequence 4SM (Fig. 2B). As a control, we verified that extracts prepared from *E. coli* containing an empty expression vector did not induce formation of complexes when used in EMSAs (results not shown).

The *Str* G-box directs seed-specific expression in transgenic tobacco plants

To investigate whether the *Str* G-box can function as a regulatory sequence, we chose a gain-of-function approach, involving the analysis of wild-type (4SW) and mutant (4SM) tetramers of the *Str* G-box in combination with two different truncated CaMV 35S promoters (Fig. 1B). The CaMV 35S −47 minimal promoter contains the sequence from −47 to +27 including the TATA box and the transcriptional start site. The slightly larger CaMV 35S –90 minimal promoter (−90 to +27) includes, in addition, the *as-1* enhancer element that is located between positions −90 and −47 (Lam et al. 1989). The −90 CaMV 35S sequence is known to act synergistically with other regulatory elements (Lam et al. 1989; Lam and Chua 1990). As shown in Fig. 3, construct 4SW-90 was active in leaves of transgenic plants, whereas the construct 4SM-90, in which the G-box tetramer is mutated, gave considerably lower expression levels. The 4SW and 4SM sequences did not give rise to detectable expression in leaves when assayed on the −47 CaMV 35S minimal promoter (Fig. 3). Since certain G-box(-like) sequences have been reported to be involved in conferring tissue-specific expression (Salinas et al. 1992; Thomas 1993), we then determined whether activity of the 4SW-47 construct could be detected in tissues other than leaves. This revealed that 4SW-47 activity was absent in roots and stem (data not shown), but was clearly detectable in seeds (Fig. 3). In contrast to the wild-type G-box construct 4SW-47, the mutant construct 4SM-47 showed no expression in seeds (Fig. 3). Similarly, construct 4SW-90 was highly active in seeds, whereas 4SM-90 gave considerably lower expression levels (Fig. 3). In summary, the 4SW tetramer, but not the mutant derivative 4SM, was found to confer expression in seeds when analysed on the −47 as well as on the −90 CaMV 35S promoter, whereas it was found to confer expression in leaves only when present on the −90 promoter. Thus, for expression in leaves, the *Str* G-box requires an additional enhancer, whereas in seeds a G-box tetramer can mediate expression independently of other known regulatory sequences.

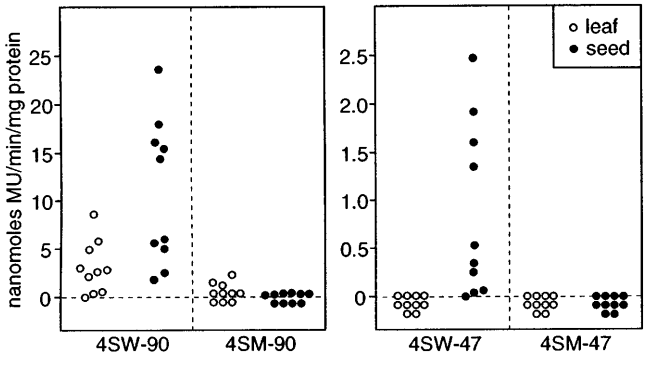

Fig. 3 *Str* G-box tetramers confer expression in tobacco leaf and seed. GUS activities were measured in protein extracts prepared from leaves (*open symbols*) or seeds (*filled symbols*) originating from 10 independent transgenic tobacco plants. Symbols *below* the zero level indicate that GUS activities were below the detection limit. Note the different scales used in the two panels

Str G-box-directed seed-specific expression is tightly controlled during plant development

To investigate at which stage of seed development the 4SW-47 construct is switched on, the expression of the wild-type and mutant constructs was followed during the development of seed capsules in mature greenhouse-grown plants, as well as during seedling development. Seed capsules of primary transformant transgenic plants were harvested at various times after flowering and were analysed for GUS activity by histochemical staining with X-Gluc as a substrate. For each construct, the results in one representative transgenic plant line are presented. In each case, three other independent transgenic plant lines showed the same expression pattern. As shown in Fig. 4A and B, at 5 and 10 days after flowering (DAF) no expression was found in seeds harbouring construct 4SW-47. At 15 DAF GUS staining was visible in seeds (Fig. 4C), whereas the surrounding tissue of the seed capsule did not stain. In seeds harbouring construct 4SM-47, GUS staining was undetectable at all three developmental stages analysed (Fig. 4G–I). These results confirm that the *Str* G-box tetramer cloned on the truncated −47 CaMV 35S promoter can confer seed-specific reporter gene expression. Moreover, this expression pattern is developmentally controlled, since it appears from 15 DAF onwards.

As shown in Fig. 4D, in seeds harbouring construct 4SW-90, GUS staining was already observed at the first time point studied (5 DAF) and the neighbouring tissues in the seed capsules were also stained. The expression pattern was similar at 10 and 15 DAF (Fig. 4E, F). Thus, expression of the 4SW-90 construct in seed capsules is not restricted to seeds and is not dependent on the developmental stage. The expression of construct 4SM-90 (Fig. 4J–L) was qualitatively similar, but considerably lower than that of the wild-type construct 4SW-90. This indicates that the wild type G-box tetramer acts as an enhancer in seeds as well as in seed capsule tissues when combined with the CaMV 35S −90 promoter. Since the plants were self-pollinated it is

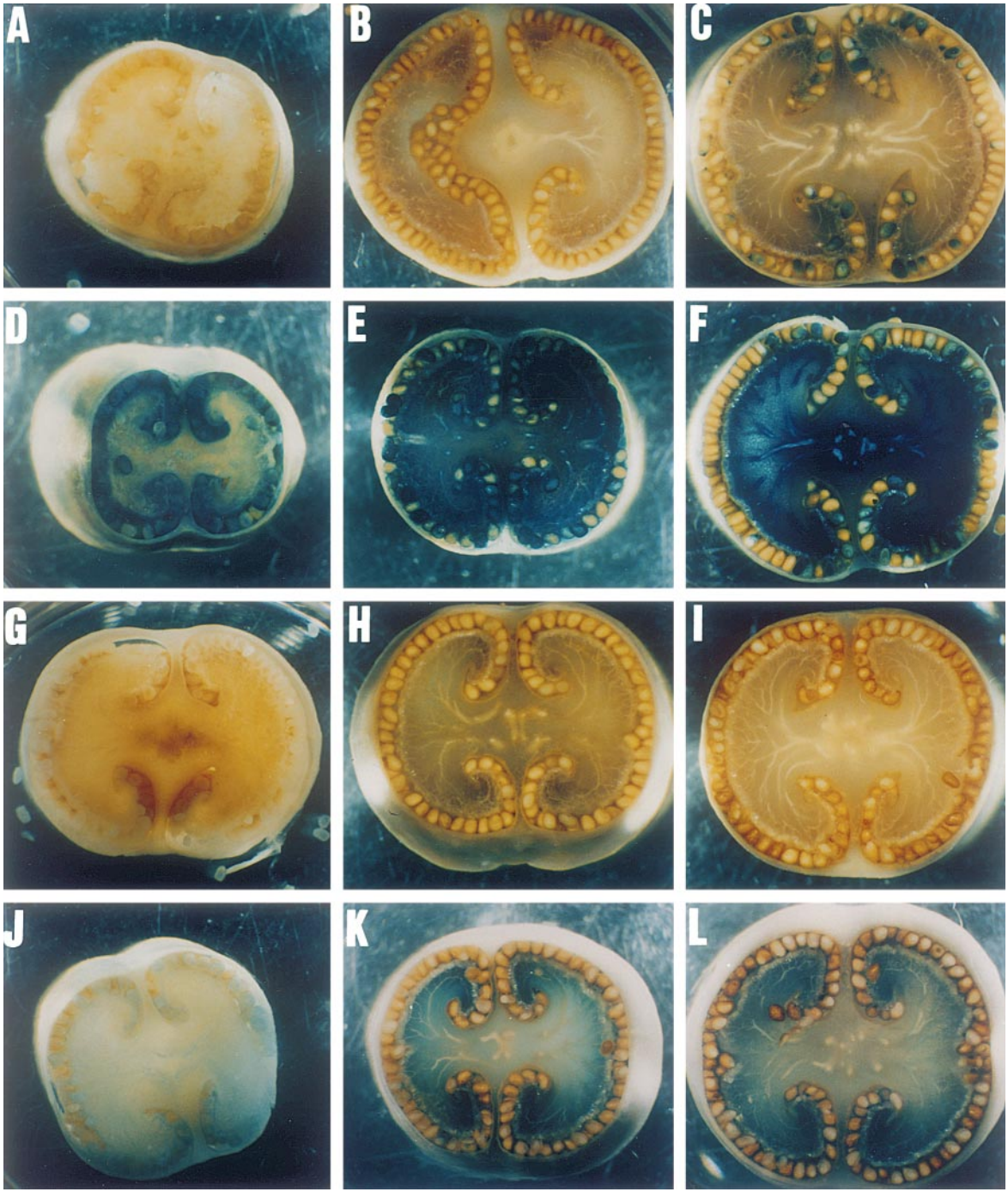

Fig. 4A–L Histochemical localization of GUS activity in tobacco seed capsules at different developmental stages. **A–C** Seed capsules containing construct 4SW-47 at 5, 10 and 15 DAF, respectively. **D–F** Seed capsules containing construct 4SW-90 at 5, 10 and 15 DAF, respectively. **G–I** Seed capsules containing construct 4SM-47 at 5, 10 and 15 DAF, respectively. **J–L** Seed capsules containing construct 4SM-90 at 5, 10 and 15 DAF, respectively

expected that at least 75% of the seeds should show GUS activity. Figure 4 shows a lower percentage of stained seeds, due to the fact that especially after 10 and 15 DAF the seed coat becomes impermeable to the X-Gluc staining solution, and therefore GUS activity can only be observed in cross-sectioned seeds.

GUS activities in developing transgenic seedlings were monitored histochemically with the substrate X-Gluc between 1 and 14 days after germination. Construct 4SW-47 showed no expression during any stage of seedling development (results not shown). Apparently, there is no expression of construct 4SW-47 after seed germination. Thus, it appears that expression of construct 4SW-47 is switched on transiently during the later stages of seed maturation and that *gusA* mRNA and GUS protein stored in the seed are rapidly degraded upon germination.

Str gene expression in *C. roseus* plants

The *Str* gene was previously shown to be expressed in different organs of *C. roseus* plants (Pasquali et al. 1992), but its expression in seeds has not been studied. To determine whether our observation that the *Str* G-box can drive seed-specific expression in gain-of-function contexts is relevant for the understanding of *Str* regulation we investigated whether the *Str* gene is indeed expressed in *C. roseus* seeds. As shown in Fig. 5, the *Str* transcript was also present in ripening seeds harvested from seed capsules 4 weeks after flowering. *Str* mRNA levels in seeds were similar to the levels in other plant organs. The presence of *Str* mRNA in seeds is consistent with the occurrence of the terpenoid indole alkaloid tabersonine in *C. roseus* seeds (De Luca et al. 1986). Tabersonine biosynthesis requires availability of the precursor strictosidine, and thus implies that *Str* is active.

Discussion

G-box elements occur in a variety of plant promoters and are reported to function in the regulation of gene expression in response to various environmental and developmental cues. In this study we focused on the role of the G-box in the promoter of the *Str* gene from *C. roseus*, which is involved in the production of a group of secondary metabolites known as terpenoid indole alkaloids. For this, two approaches were followed, which involved gain-of-function studies with chimaeric gene constructs in transgenic tobacco plants, and the in vitro analysis of G-box binding protein factors.

Using EMSAs it was demonstrated that the *Str* G-box is bound in vitro by factors present in tobacco nuclear extracts. Since not all G-box-containing DNA fragments are recognized by such factors in vitro (Williams et al. 1992; Izawa et al. 1993), this observation is a first indication that the G-box in the context of the *Str* promoter may indeed represent a functional element. The shifted complex of the *Str* G-box was diffuse and of low mobility, which is similar to the behaviour of complexes formed by binding of nuclear factors to G-box containing sequences from the promoters of several plant genes (Giuliano et al. 1988; McKendree et al. 1990; Schindler et al. 1992; Hong et al. 1995). Such characteristic diffuse complexes of low or high mobility (type A or B complexes, Williams et al. 1992) have also been observed with a number of different bZIP transcription factors that interact with certain G-box sequences (Izawa et al. 1993). This may suggest that bZIP proteins in our nuclear extracts produce the *Str* G-box shift. In agreement with this idea, we found that a similar low-mobility complex was obtained when the *Str* G-box was incubated with the tobacco G-box binding bZIP factor TAF-1, synthesized in *E. coli*, instead of with leaf nuclear extract. The specificity of the complexes was confirmed by demonstrating that a mutant *Str* G-box sequence, in which two nucleotides had been changed, showed reduced binding of tobacco nuclear factors and of TAF-1.

Fig. 5 Expression of *Str* in different tissues of mature *C. roseus* plants. A gel blot containing total RNA from flowers (F), leaves (L), stems (St), roots (R) and developing seeds (S), harvested 4 weeks after flowering, was probed with the *Str* cDNA (Pasquali et al. 1992)

To investigate the function of this G-box further, we analysed the activity of wild-type (4SW) and mutant (4SM) *Str* G-box tetramers combined with CaMV 35S promoters truncated at position −47 or −90. In leaves, GUS activities obtained with the construct 4SW-90 were considerably higher than those seen with the mutant construct 4SM-90. Thus, the two single-nucleotide mutations which disrupt the palindromic structure of the *Str* G-box and thereby reduce binding of nuclear factors in vitro, also reduce transcription-enhancing activity in vivo. In leaves, the G-box tetramer was found to confer expression only on the −90 but not on the −47 CaMV 35S promoter. Therefore, it appears that the G-box acts synergistically with the CaMV 35S promoter sequence between positions −90 and −47. This is probably due to interaction with the *as-1* enhancer in the −90 to −47 region, which is known bind bZIP proteins (Katagiri et al. 1989; Lam et al. 1989). It is quite possible that the G-box also synergizes with other *cis*-acting elements in the context of the native *Str* promoter. Such synergism has also been reported for G-box elements in other promoter contexts (Donald and Cashmore 1990; Kawagoe et al. 1994; Loake et al. 1992; Shen and Ho 1995).

In *C. roseus* the *Str* gene is expressed in various plants organs (Pasquali et al. 1992), and in developing seeds as shown here. Our studies on the tissue-specific expression of the G-box constructs revealed that construct 4SW-47 did not drive expression in the different tissues of seedlings tested or in mature plants, but was active at later stages of seed development and in mature seeds. The −47 to +27 CaMV 35S promoter sequence has been used frequently for promoter studies and it has never been reported to contain any *cis*-acting sequences other than the TATA box. Thus, it is very likely that the seed-specific expression of construct 4SW-47 results from the activity of the *Str* G-box alone. This is also supported by the fact that the two single-nucleotide mutations completely eliminate expression. Apparently, the *Str* G-box element confers developmentally controlled seed-specific expression as an autonomous *cis*-acting element. Just as in leaves, analysis of the 4SW-90 and 4SM-90 constructs indicated that the G-box can also act synergistically with the −90 to −47 region to enhance gene expression in seeds at early developmental stages and in seed capsule

tissues. Autonomous action of a true G-box sequence (CACGTG), such as that seen with the 4SW-47 construct, has not been reported in other studies of G-box elements using stably transformed plants. Up to now it has only been shown in a transient expression study with protoplasts that a tetramer of a G-box element could confer abscisic acid responsiveness on a reporter gene when assayed in combination with the truncated −40 CaMV 35S promoter (Vasil et al. 1995). To the best of our knowledge, in all other studies reported to date (e.g. Oeda et al. 1991; Loake et al. 1992; Salinas et al. 1992; Kawagoe et al. 1994; Shen and Ho 1995) other known regulatory sequences were present and in all likelihood synergized with the G-box.

Seed-specific expression has also been reported for a number of elements that resemble the true G-box (CACGTG), for example the artificial *hex3* element (containing the sequence CGCGTG; Lam and Chua 1991), the *Iwt* element (GGTACGTGGC) found in various abscisic acid-responsive genes (Marcotte et al. 1989; Mundy et al. 1990; Salinas et al. 1992), and the W1 sequence (containing the G-box-like sequence TACGTG, and an odd-base C-box, ATGAGTCAT) from the pea lectin promoter (De Pater et al. 1993). Possible candidates for the *trans*-acting factors that mediate seed-specific *Str* G-box expression in tobacco are tobacco homologues of the bZIP proteins Opaque-2 (O2) from maize (Hartings et al. 1989) or RITA-1 from rice (Izawa et al. 1994). Both O2 and RITA-1 have been shown to be capable of binding to G-box sequences and are specifically expressed in seeds (Izawa et al. 1993, 1994; Ueda et al. 1992; Varagona et al. 1991). Alternatively, the G-box may be a target for MYC-related bHLH proteins (De Pater et al. 1997). TAF-1, which was shown here to bind to the *Str* G-box in vitro, is unlikely to be involved in the seed-specific expression, since its mRNA is not detectable in seeds (Oeda et al. 1991).

What determines whether or not a G-box(-like) element confers seed-specific expression is an intriguing question. The mutational analysis of the *Str* G-box showed that the integrity of the palindromic structure is required for expression in seeds. However, not all promoters that contain a true palindromic G-box sequence are expressed in seeds. For example, the G-box-containing sequence PA (GCCACGTGGC) tested as a tetramer on a −90 CaMV 35S truncated promoter gave enhanced expression in leaves and roots, but not in seeds (Salinas et al. 1992). The differences in function between PA and the *Str* G-box are probably due to differences in the nucleotides flanking the G-box. It has been reported that the affinity of TAF-1 and other bZIP factors for G-box-containing elements is strongly influenced by the nucleotides upstream as well as downstream of the G-box core sequence (Izawa et al. 1993). Consequently this will affect the expression patterns of genes containing different G-box-flanking sequences. The PA and *Str* G-box sequences, like the G-box-like sequence *Iwt* (GGTACGTGGC) that confers seed-specific expression, were all analysed on the −90 CaMV 35S promoter which enables a direct comparison of their characteristics. In Table 1 an alignment of the PA, *Str* G-box, and *Iwt* sequences is shown. The nucleotides are numbered following the nomenclature of Williams et al. (1992), in which the central nucleotides C and G are designated as −0 and +0 respectively. All three sequences share the sequence ACGTGG from −1 to +3, and PA and the *Str* G-box differ only in positions −4 and +4. As PA does not confer expression in seeds, whereas the *Str* G-box does, positions −4 and/or +4 may determine seed specificity. However, these cannot be the only determinants since they are identical in PA and *Iwt*, whereas only *Iwt* drives expression in seeds. Therefore, positions −3 and −2, which differ between PA and *Iwt*, may also determine seed specificity. In conclusion, the G-box or G-box-like core sequence is not sufficient for seed specificity, but nucleotides at positions −4, −3, −2 and/or +4 may be additional determinants.

Apart from their role in conferring tissue-specific expression patterns, G-box(-like) elements have also been implicated in mediating reponses to environmental signals. In *C. roseus*, *Str* expression was shown to be inducible by fungal elicitors (Pasquali et al. 1992) and UV light (our unpublished results) and it was also demonstrated that a reporter gene construct driven by the *Str* promoter can be induced by treatment with yeast extract elicitor in tobacco (Pasquali et al. 1999). Since other studies have suggested the involvement of G-box sequences in responses to UV light (Schulze-Lefert et al. 1989a) or fungal elicitor (Dröge-Laser et al. 1997), we examined the effect of such signals on expression of the different *Str* G-box constructs. Neither yeast extract elicitor nor UV light was found to induce their expression, whereas these factors did induce the activity of constructs containing either the native *Str* promoter or promoters of other genes required for biosynthesis of terpenoid indole alkaloids (results not shown). Therefore, it appears that the *Str* G-box does not confer responsiveness to elicitor or UV light. These observations

Table 1 Sequence comparison of PA, *Iwt* and the G-box-containing element from the *Str* promoter

Cis-acting elements	Nucleotide sequence										Expression in seeds	Reference
	−4	−3	−2	−1	−0	+0	+1	+2	+3	+4		
PA	G	C	C	A	C	G	T	G	G	C	–	Salinas et al. (1992)
Iwt	G	G	T	A	C	G	T	G	G	C	+	Salinas et al. 1992)
Str	T	C	C	A	C	G	T	G	G	T	+	This work

are confirmed by the fact that mutation of the G-box in the native *Str* promoter context does not affect elicitor responsiveness (Pasquali et al. 1999).

In summary, the results of this study show that binding of nuclear factors and TAF-1 to the wild-type *Str* G-box in vitro is reduced by mutations which also decrease activity in vivo. We therefore propose that, in the *Str* promoter context, the G-box is a functional *cis*-acting element. The results of gain-of-function experiments suggest that this G-box is involved in mediating *Str* expression in seeds and may also contribute to the *Str* expression level in other tissues by acting synergistically with other regulatory elements. An understanding of the *cis*-acting elements and *trans*-acting factors involved in regulation of the expression of *Str* and other genes required for terpenoid indole alkaloid biosynthesis may eventually lead to strategies for genetic modification of terpenoid indole alkaloid levels in *C. roseus*.

Acknowledgements The authors wish to thank Dr. B. S. de Pater for providing TAF-1 extract, B. Wetselaar for taking care of plants, P. Hock for preparation of figures, W. Snoeijer from the Department of Pharmacognosy (Gorlaeus Laboratories) for the *C. roseus* plants and Dr. A. H. Meijer for technical advice and valuable discussions.

References

Aerts RJ, Snoeijer W, Aerts-Teerlink O, Van der Meijden E, Verpoorte R (1991) Control and biological implications of alkaloid synthesis in *Cinchona* seedlings. Phytochemistry 30:3571–3577

Block A, Dangl JL, Hahlbrock K, Schulze-Lefert P (1990) Functional borders, genetic fine structure, and distance requirements of *cis* elements mediating light responsiveness of the parsley chalcone synthase promoter. Proc Natl Acad Sci USA 87:5387–5391

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Chattopadhyay S, Ang LH, Puente P, Deng XW, Wei N (1998) *Arabidopsis* bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. Plant Cell 10:673–683

Chomczynsky P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

De Luca V, Balsevich J, Tyler RT, Eilert U, Panchuk BD, Kurz WGW (1986) Biosynthesis of indole alkaloids: Developmental regulation of the biosynthetic pathway from tabersonine to vindoline in *Catharanthus roseus*. J Plant Physiol 125:147–156

De Pater S, Pham K, Chua N–H, Memelink J, Kijne J (1993) A 22-bp fragment of the pea lectin promoter containing essential TGAC-like motifs confers seed-specific gene expression. Plant Cell 5:877–886

De Pater S, Pham K, Memelink J, Kijne JW (1997) RAP-1 is an Arabidopsis MYC-like R protein homologue, that binds to G-box sequence motifs. Plant Mol Biol 34:169–174

Dolferus R, Jacobs M, Peacock WJ, Dennis ES (1994) Differential interactions of promoter elements in stress responses of the *Arabidopsis Adh* gene. Plant Physiol 105:1075–1087

Donald RGK, Cashmore AR (1990) Mutation of either G box or I box sequences profoundly affects expression from the *Arabidopsis rbcS-1A* promoter. EMBO J 9:1717–1726

Dröge-Laser W, Kaiser A, Lindsay WP, Halkier BA, Loake GJ, Doerner P, Dixon RA, Lamb C (1997) Rapid stimulation of a soybean protein-serine kinase that phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses. EMBO J 16:726–738

Foster R, Izawa T, Chua N–H (1994) Plant bZip proteins gather at ACGT elements. FASEB J 8:192–200

Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik PA, Cashmore AR (1988) An evolutionary conserved protein binding sequence upstream of a plant light-regulated gene. Proc Natl Acad Sci USA 85:7089–7093

Guiltinan MJ, Marcotte WR, Quatrano RS (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. Science 250:267–270

Green PJ, Kay SA, Lam E, Chua N–H (1991) *In vitro* DNA footprinting. In: Gelvin SB, Schilperoort RA, Verma DPS (eds) Plant Molecular Biology Manual. Kluwer Academic Publishers, Dordrecht, B11: pp 1–22

Hartmann T (1991) Alkaloids. In: Rosenthal GA, Berenbaum MR (eds) Herbivores: Their Interaction with Secondary Plant Metabolites, Academic Press, San Diego, Vol. 1. pp 79–121

Hartings H, Maddaloni M, Lazzaroni N, Di Fonzo N, Motto M, Salamini F, Thompson R (1989) The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. EMBO J 8:2795–2801

Hashimoto T, Yamada Y (1994) Alkaloid biogenesis: Molecular aspects. Annu Rev Plant Physiol Plant Mol Biol 45:257–285

Hong JC, Cheong YH, Nagao RT, Bahk JD, Key JL, Cho MJ (1995) Isolation of two soybean G-box binding factors which interact with a G-box sequence of an auxin-responsive gene. Plant J 8:199–211

Horsch RB, Fry J, Hoffmann N, Neidermeyer J, Rogers SG, Fraley RT (1991) Leaf disc transformation. In: Plant Molecular Biology Manual. In: Gelvin SB, Schilperoort RA, Verma DPS (eds) Plant molecular biology manual. Kluwer Academic Publishers, Dordrecht, A5:pp 1–9

Izawa T, Foster R, Chua N–H (1993) Plant bZIP protein DNA binding specificity. J Mol Biol 230:1131–1144

Izawa T, Foster R, Nakajima M, Shimamoto K, Chua N–H (1994) The rice bZip transcriptional activator RITA-1 is highly expressed during seed development. Plant Cell 6:1277–1287

Jefferson RA (1987) Assaying chimeric genes in plants: The GUS gene fusion system. Plant Mol Biol Rep 5:387–405

Kawagoe Y, Campell BR, Murai N (1994) Synergism between CACGTG (G-box) and CACCTG *cis*-elements is required for activation of the bean seed storage protein β-phaseoline gene. Plant J 5:885–890

Katagiri F, Lam E, Chua N–H (1989) Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature 340:727–730

Lam E, Benfey PN, Gilmartin PM, Fang R–X, Chua N–H (1989) Site-specific mutations alter in vitro factor binding and change promoter expression pattern in transgenic plants. Proc Natl Acad Sci USA 86:7890–7894

Lam E, Chua N-H (1990) GT–1 binding site confers light responsive expression in transgenic tobacco. Science 248:471–473

Lam E, Chua N–H (1991) Tetramer of a 21-base pair synthetic element confers seed expression and transcriptional enhancement in response to water stress and abscisic acid. J Biol Chem 266:17131–17135

Loake GJ, Faktor O, Lamb CJ, Dixon RA (1992) Combination of H-box [CCTACC(N)7CT] and G-box (CACGTG) *cis* elements is necessary for feed-forward stimulation of a chalcone synthase promoter by the phenylpropanoid-pathway intermediate *p*–coumaric acid. Proc Natl Acad Sci USA 89:9230–9234

Luijendijk TJC (1995) Strictosidine glucosidase in indole alkaloid biosynthesis. Ph. D. thesis, Leiden University, The Netherlands

Marcotte WR Jr, Russell SH, Quatrono RS (1989) Abscisic acid-responsive sequences from the Em gene of wheat. Plant Cell 1:969–976

McKendree WL, Paul A-L, DeLisle AJ, Ferl RJ (1990) In vivo and in vitro characterization of protein interactions with the dyad G-box of the *Arabidopsis Adh* gene. Plant Cell 2:207–214

McKnight TD, Roessner CA, Devagupta R, Scott AI, Nessler CL (1990) Nucleotide sequence of a cDNA encoding the vacuolar protein strictosidine synthase from *Catharanthus roseus*. Nucl. Acids Res 18:4939

Meier I, Gruissem W (1994) Novel conserved sequence motifs in plant G-box binding proteins and implications for interactive domains. Nucl Acids Res 22:470–478

Meijer AH, Verpoorte R, Hoge JHC (1993) Regulation of enzymes and genes involved in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. J Plant Res 3:145–164

Memelink J, Swords KMM, Staehelin LA, Hoge JHC (1994) Southern, Northern and Western blot analysis. In: Gelvin SB, Schilperoort RA, Verma DPS (eds) Plant Molecular Biology Manual Kluwer Academic Publishers, Dordrecht, F1:pp 1–23

Menkens AE, Cashmore AR (1994) Isolation and characterization of a fourth *Arabidopsis thaliana* G-box binding factor, which has similarity to Fos oncoprotein. Proc Natl Acad Sci USA 91:2522–2526

Menkens AE, Schindler U, Cashmore AR (1995) The G-box: a ubiquitous regulatory element in plants bound by the GBF family of bZip proteins. Trends Biochem Sci 20:506–510

Meshi T, Iwabuchi M (1995) Plant transcription factors. Plant Cell Physiol 36:1405–1420

Mundy J, Yamaguchi-Shinozaki K, Chua N–H (1990) Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice *rab* gene. Proc Natl Acad Sci USA 87:1406–1410

Oeda K, Salinas J, Chua N-H (1991) A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J 7:1793–1802

Ouwerkerk PBF, Memelink, J (1997) A simple method for directional multimerization of DNA sequences. Trends Genet 13:207

Pasquali G, Goddijn OJM, de Waal A, Verpoorte R, Schilperoort RA, Hoge JHC, Memelink J (1992) Coordinated regulation of two indole alkaloid biosynthetic genes from *Catharanthus roseus* by auxin and elicitors. Plant Mol Biol 18:1121–1131

Pasquali G, Ouwerkerk PBF, Memelink J (1994) Versatile transformation vectors to assay the promoter activity of DNA elements in plants. Gene 149:373–377

Pasquali G, Erven ASW, Ouwerkerk PBF, Menken FLH, Memelink J (1999) The promoter of the strictosidine synthase gene from periwinkle confers elicitor-inducible expression in transgenic tobacco and binds nuclear factors. Plant Mol Biol, in press

Salinas J, Oeda K, Chua N-H (1992) Two G-box-related sequences confer different expression patterns in transgenic tobacco. Plant Cell 4:1485–1493

Schindler U, Menkens AE, Beckmann H, Ecker JR,Cashmore AR (1992) Heterodimerization between light-regulated and ubiquitously expressed *Arabidopsis* GBF bZIP proteins. EMBO J 11:1261–1273

Schulze-Lefert P, Dangl JL, Becker-Andre M, Hahlbrock K, Schulz W (1989a) Inducible in vivo footpints define sequences necessary for UV-light activation of the parsley chalcone synthase gene. EMBO J 8:651–656

Schulze-Lefert P, Becker-Andre M, Schulz W, Hahlbrock K, Dangl JL (1989b) Functional architecture of the light-responsive chalcone synthase promoter from parsley. Plant Cell 1:707–714

Shen Q, Ho T–H D (1995) Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel *cis*-acting element. Plant Cell 7:295–307

Thomas TL (1993) Gene expression during plant embryogenesis and germination: An overview. Plant Cell 5:1401–1410

Thomas JC, Adams DG, Nessler CL, Brown JK, Bohmert HJ (1995) Tryptophan decarboxylase, tryptamine, and reproduction of the whitefly. Plant Physiol 109:717–720

Ueda T, Waverczak W, Ward K, Sher N, Ketufat M, Schmidt RJ, Messing J (1992) Mutations of the 22- and 27-kD zein promoters affect transactivation by the Opaque–2 protein. Plant Cell 4:701–709

Urwin NA, Jenkins GI (1997) A sucrose repression element in the *Phaseolus vulgaris rbcS2* gene promoter resembles elements responsible for sugar stimulation of plant and mammalian genes. Plant Mol Biol 35:929–942

Van Slogteren GMS, Hoge JHC, Hooykaas PJJ, Schilperoort RA (1983) Clonal analysis of heterogenous crown gall tumor tissues induced by wild-type and shooter mutant strains of *Agrobacterium tumefaciens:* expression of T–DNA genes. Plant Mol Biol 2:321–333

Varagona MJ, Schmidt RJ, Raikhel NV (1991) Monocot regulatory protein Opaque–2 is localized in the nucleus of maize endosperm and transformed tobacco plants. Plant Cell 3:105–113

Vasil V, Marcotte Jr WR, Rosenkrans L, Cocciolone SM, Vasil IK, Quatrono RS, McCarty DR (1995) Overlap of Viviparous (VP1) and abscisic acid response elements in the *Em* promoter: G-box elements are sufficient but not necessary for VP1 transactivation. Plant Cell 7:1511–1518

Weisshaar B, Armstrong GA, Block A, da Costa e Silva O, Hahlbrock K (1991) Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. EMBO J 10:1777–1786

Williams ME, Foster R, Chua N–H (1992) Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. Plant Cell 4:485–496