

Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in *Arabidopsis thaliana* cycling cells

Dominique Trémousaygue, Lionel Garnier, Claude Bardet, Patrick Dabos, Christine Hervé and Bernard Lescure*

Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27, 31326 Castanet-Tolosan, France

Received 4 October 2002; accepted 14 November 2002.

*For correspondence (fax +33 0561285061; e-mail lescure@toulouse.inra.fr).

Summary

We have focused our interest on two *cis*-regulatory elements, named site II motif and *telo* box, identified within the promoter of plant *proliferating cellular nuclear antigen* (*PCNA*) and putatively involved in meristematic expression of the gene. A conserved topological association between site II motifs and *telo* boxes is observed in the promoter of numerous genes expressed in cycling cells, including several cell cycle-related genes and 153 *Arabidopsis* genes encoding ribosomal proteins. Meristematic expression of a GUS reporter gene was observed in plants under the control of *Arabidopsis* site II motif within a minimal promoter. This expression is strongly enhanced by addition of a *telo* box within this chimaeric promoter. We showed by gel retardation experiments that the site II motif is a target for several DNA-binding activities present in *Arabidopsis* crude cell extract and can bind a transcription factor, At-TCP20, from the Teosinte branched 1, Cycloidea, PCF (TCP)-domain protein family. In yeast two-hybrid experiments, At-TCP20 appears to be a potential partner of AtPur α , which was previously shown to bind *telo* boxes. An important consequence of this analysis is to reveal new and conserved regulatory processes concerning the regulation of plant ribosomal gene expression in cycling cells. The implication of these observations in plant-specific developmental pathways is discussed.

Keywords: cell division, ribosomal protein gene, *PCNA*, transcription regulation, AtPur α , At-TCP20.

Introduction

Whereas animals have a defined and early embryonic phase in which all of the major organ systems develop, plants maintain a life-long capacity for morphogenesis. Except for the embryonic stage of growth, plant morphogenesis essentially results from the progressive development and activity of meristems during plant growth. Cell cycle progression within the meristematic area is controlled by key processes which share common features with animals. A major control point occurs during the G₁-S transition and is mainly exerted at the transcriptional level by conserved molecular processes involving cyclins and Cyclin dependant kinases (CDKs), as well as pRB, and E2F recently identified in plants (reviewed by Meijer and Murray, 2001). Nutrient and metabolic capacity limitations interfere with these general control processes and are critical for cell cycle progression. For example, sugar levels that reflect the overall photosynthetic capacity of a plant regulate the expression of cyclin D (De Veylder *et al.*, 1999;

Riou-Khamlichi *et al.*, 2000). As in other eukaryotes, the rate of protein synthesis appears to be a critical determinant of cellular growth (Ohnishi *et al.*, 1990). Strong stimulation of the expression of genes encoding components of the translational apparatus is observed in cycling cells and upon the addition of the plant hormone auxin (Axelos *et al.*, 1992; Gantt and Key, 1985; Gao *et al.*, 1994). Conversely, mutations in several genes encoding ribosomal proteins (RP) in *Arabidopsis* result in morphological alterations and a significant delay in development (Ito *et al.*, 2000; Van Lijsebetens *et al.*, 1994; Weijers *et al.*, 2001). Studies of promoters of plant genes encoding components of the translational machinery have led to the identification of several regulatory elements that participate in the over-expression of RP genes in dividing cells. The *telo* box (AAACCCTAA) was first observed within the four promoters of *Arabidopsis* genes encoding the translation elongation factor EF1A and subsequently within several plant RP promoters (Axelos *et al.*,

1989; Liboz *et al.*, 1990; Regad *et al.*, 1995; Tremoussaygue *et al.*, 1999), and it is identical to the (AAACCCT)_n repeat motif of plant telomeres (Richards and Ausubel, 1988). This observation is reminiscent of the characteristic features of yeast genes encoding components of the translational machinery. The promoters of yeast *EF1A* and *RP* genes contain a regulatory element, the *rpg*-box, which shows structural similarities with the telomeric repeat of yeast and interacts with RAP1, a protein which has pleiotropic functions, participating both in the regulation of expression of house-keeping genes and in controlling size and stability of yeast telomeres (for reviews see Shore, 1994, 1997). In spite of this structural similarity, a mutation in the *telo* box within the full-length *Arabidopsis EF1A A1* promoter failed to affect its constitutive activity (Curie *et al.*, 1991, 1993). Nevertheless, we suggested that this element could be involved in only one of the several superimposed plant developmental programmes controlling expression of the *EF1A A1* gene, and therefore the effect of its deletion could be masked by molecular activation processes not affected by this element. Recent reported data suggest that this is indeed the case. In gain-of-function experiments, the *telo* box, in synergy with other *cis*-acting elements, the *tef* or the *trap* box, participates in the control of gene expression in root meristems (Tremoussaygue *et al.*, 1999). The *tef* box (ARGGRYANNNNGT) was initially characterised within the promoters of *Arabidopsis* genes encoding the translation elongation factor *EF1A A1* (Curie *et al.*, 1991, 1993) and several *Arabidopsis RP* genes (Regad *et al.*, 1995). The *trap40* box (GGGGGTAGAATAG) was identified within the promoter of an *Arabidopsis* gene encoding an acidic ribosomal protein (Scheer *et al.*, 1997). The *tef* box is the target for two heteromeric protein complexes, C1 and C2 (Manevski *et al.*, 1999), and as C1 also interacts with the *trap40* box, functional relationships exist between these elements (Manevski *et al.*, 1999). By screening an expression cDNA library with a double-stranded plant telomere motif, an *Arabidopsis* gene encoding a protein that can interact with the *telo* boxes of plant *EF1A* and *RP* promoters has been isolated (Tremoussaygue *et al.*, 1999). This protein, named AtPur α , is related to the animal Pur α protein, first identified in humans (Bergemann *et al.*, 1992). Interestingly, *telo* boxes are also observed within the promoter of other plant genes expressed in late G₁, such as *ribonucleotide reductase (RNR)* or *proliferating cell nuclear antigen (PCNA)*, raising the possibility that this element could be involved in a common regulatory process which connects expression of a set of genes at the G₁-S transition. In addition to a *telo* box located downstream of the transcription initiation site, the rice *PCNA* promoter contains two *cis*-acting elements (the site *Ila* and *Ilb* motifs) which regulate gene expression in meristems of transgenic tobaccos and specifically interact with two rice bHLH transcription factors (PCF1 and PCF2) (Kosugi and Ohashi, 1997, 2002; Kosugi

et al., 1995). Moreover, these rice elements activate the expression in *Arabidopsis* root primordia in synergy with a *telo* box (Manevski *et al.*, 2000). These data suggest that site II motifs participate in a conserved regulatory process acting in diverse plant species.

In this report, we characterise a *cis*-acting element of the *Arabidopsis PCNA-2* gene promoter showing the same transcriptional activation properties as site *Ila* and *Ilb* rice *cis*-acting elements. It is the target for at least five DNA-binding activities of cellular protein extracts prepared from exponentially growing *Arabidopsis* cell suspension cultures. It specifically binds to an *Arabidopsis* protein (At-TCP20) related to PCF1 and 2. These factors belong to a family of plant-specific proteins sharing a common motif called the 'TCP' domain (Cubas, 2002; Cubas *et al.*, 1999). All members of the TCP gene family investigated so far appear to play a role in processes related to plant development (Cubas, 2002). Moreover, we demonstrate a physical interaction between At-TCP20 and AtPur α . Strikingly, as for plant *PCNA* gene promoters, a conserved topological association of the *Arabidopsis cis*-acting element II and a *telo* box is found within the majority of *Arabidopsis* genes encoding ribosomal proteins. The implications of these observations are discussed.

Results

The site II motifs of the Arabidopsis PCNA-2 promoter are sufficient to activate gene expression in cycling cells and act in synergy with a telo box

Rice *PCNA* gene promoter sequence analysis led to the identification of regulatory elements *Ila* and *Ilb* which are essential for the proliferating cell-specific transcriptional promoter activity (Kosugi *et al.*, 1995). The *Arabidopsis* genome harbours two genes encoding PCNA located on chromosomes I and II, respectively. The analysis of the two corresponding promoters indicated that the gene of chromosome II (*PCNA-2*) has two GC-rich duplicated motifs (TTGGGCC) similar to the *Ila* element of the rice *PCNA* promoter (TGGGCC) (Figure 1a). In addition, the *At-PCNA-2* promoter harbours two *telo* boxes (ACCCTAAA and AAACCCTAA), as already observed in the rice *PCNA* promoter (Manevski *et al.*, 2000; Figure 1a).

Here, we analyse the ability of putative *Arabidopsis PCNA-2* site II motifs to activate the expression of a GUS reporter gene in primordia of transgenic *Arabidopsis* and the effect of *telo* box on this expression. A 33-bp DNA fragment containing the two putative site II motifs of the *Arabidopsis PCNA-2* promoter was inserted, with or without a *telo* box, upstream of an *Arabidopsis EF1A A1* minimal promoter as shown in Figure 1(b). The resulting chimaeric promoters were fused upstream of the GUS

reporter gene, inserted in a binary vector and used to obtain transgenic *Arabidopsis* plants. For each construct, at least seven independent transgenic plants were analysed for the expression of GUS in F₂ young seedlings.

As illustrated in Figure 1(c, section 1), the upstream insertion of a 33-bp *Arabidopsis* PCNA-2 DNA fragment containing the two putative site II motifs led to a detectable expression of the reporter gene which was restricted to root primordia and, to a lesser extent, to young leaves. The insertion of a *telo* box within this chimaeric promoter strongly stimulated the expression of GUS, both in root primordia and in young leaves (Figure 1c, section 2). To confirm that the putative *Arabidopsis* PCNA-2 site II motifs were indeed involved in the observed gene activation, the effects of mutations on expression in these two site II motifs

were analysed. As illustrated in Figure 1(c, section 3), these mutations totally abolish the expression of the reporter gene. Thus, the putative *Arabidopsis* PCNA-2 site II motifs function in a similar way to the *Ila* and *Ilb* elements of the rice PCNA gene promoter (Kosugi *et al.*, 1995). Moreover, these observations confirm that the *telo* box increases the activation driven by various *cis*-acting elements but is not sufficient by itself to induce gene expression in cycling cells.

Association and topological location of *telo* boxes and site II motifs within the promoters of *Arabidopsis* genes encoding ribosomal proteins

As the complete DNA sequence of several eukaryotic genomes has now been determined, computational sequence

Figure 1. Effect of the *Arabidopsis* PCNA-2 *cis*-acting element II and of the *telo* box on the expression of a reporter gene.

(a) Analysis of the *Arabidopsis* PCNA-2 promoter. The putative element II (TGGGCC), *telo* boxes (AACCTAA) are coloured. The TATA box is capitalised.

(b) Schematic representation of promoter used to drive GUS gene expression. The hatched box represents the structure of the minimum *Arabidopsis* *eEF1A* promoter used for constructs (Tremousaygue *et al.* 1999). *Telo* box and *AtPCNA-2* sequences, wild type or mutated, were introduced at -48 and -77, respectively, relative to the GUS translation initiation site.

(c) Representative illustration of histochemical analysis of GUS activity observed in transgenic *Arabidopsis*. *Cis*-acting elements introduced in each type of promoter are indicated below the illustrations.

(a)

```

II           II
TTACAGCTGCTGTGGGCCGAATGTTTGTGGGCCAATTTAAACCCATGAAATAAGAAAGCTT
TTAAAGTGATGGTGGTGTACTGAAATAAACAAATACGGACTCACACAGATGATAATTCATCGTCA
CACACAGCGGGAACAAAAATACCCATAAATCGCTAAATGGCGCCACAAAATCTCAGCTTATAA
TATCACTTCCCCTAACAAATGCAATCTCCTCCAGAAATTTCAACAGCTCTAAAACCCCTAAC
CGTTTCGATTACACTCTAACCTTTCGATAATTTCCAGAAAGATG

```

telo-box

telo-box

(b)

```

wtcc 5' GCTGTTGGGCCGA
      ATGTTTGTGGCCCAATTTA 3'
OR
mut2 5' GCTGTTGAACCGA
      ATGTTTGTGGCAATTTA 3'

```

-77

```

telowt 5' TGCAGAAACCCTAACTCG 3'
OR
telomut 5' TGCAGAATTCGAAGCTCG 3'

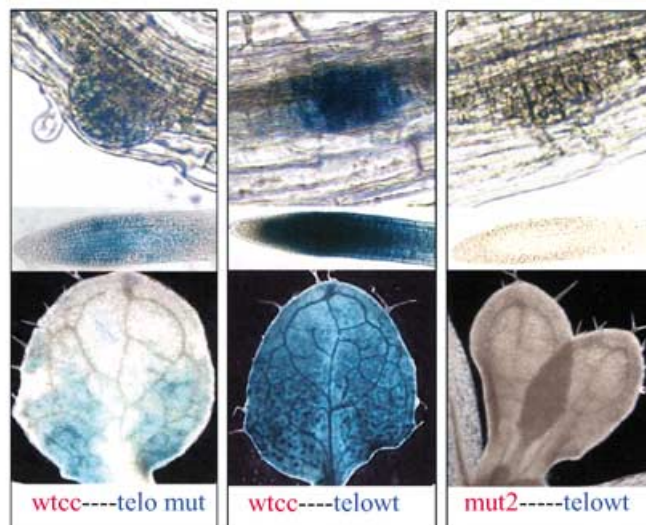
```

-48 -43

ATG

GUS

(c)



1

2

3

analysis provides new possibilities for scanning the genome to identify those regions predicted to participate in gene regulation (Pennacchio and Rubin, 2001). For example, inter-species sequence comparisons of homologous mammalian genes have been used to identify non-coding sequences which probably have regulatory functions. Non-coding sequences of co-regulated genes within a given species can also be compared. Co-expression of genes most likely reflects regulation by common transcription factors, and intra-species sequence comparison of non-coding DNA near co-expressed genes is likely to be a fruitful strategy for identifying common sequences that are important in co-ordinated gene regulation. We used this strategy to compare the promoter regions of *Arabidopsis* genes encoding ribosomal proteins. Like *PCNA* genes, ribosomal genes are over-expressed in a co-ordinated manner in cycling cells. Moreover, like plant *PCNA* promoters, many plant ribosomal promoters harbour a *telo* box which acts in synergy with various *cis*-acting elements to activate expression in cycling cells of root primordia in *Arabidopsis* (Manevski *et al.*, 2000; Trémoussaygue *et al.*, 1999). A search for the core motif of both the *telo* box (ACCCTA) and the *PCNA* site II motif (TGGGCC/T) was undertaken within the region spanning 500-bp upstream of the start codon of *Arabidopsis RP* genes using the tools available in the MIPS *Arabidopsis thaliana* Database (<http://www.mips.biochem.mpg.de/proj/thal/db/index.html>). Out of 216 promoters of genes annotated as encoding a ribosomal protein of the 40 or the 60S ribosomal subunits, 174 contained at least one *telo* box and in 153 cases this box was associated with one or several site II motifs (Figure 2a). Both TGGGCC and TGGGCT motifs were found frequently within the promoter of *RP* genes. This observation is highly significant because when the same search was conducted on a sample of 200 promoters of randomly chosen genes on the five chromosomes, only 35 of them harboured a *telo* box and seven were associated with a site II motif. Moreover, analysis of the *telo* box and site II motif locations within the *RP* promoters showed strikingly conserved topological organisation of these two elements. The majority of *telo* boxes were observed between 60- and 150-bp upstream of the translation initiation codon, whereas the site II motifs were usually located between 120- and 180-bp upstream of the start codon (Figure 2). Individual analysis of the 153 *RP* promoters containing both one or several *telo* boxes and site II motifs indicated that, except in eight cases, the *telo* box is

always located downstream from the site II motif. The compilation of nucleotides flanking the *telo* boxes and site II core motifs used for the search on ribosomal promoters (ACCCTA and TGGGCC/T, respectively) allowed a consensus sequence for each one of these elements to be defined: AAACCCTA and WWWTGGGC/TWWWWW (where W indicates either A or T) for the *telo* box and the II element, respectively (not shown). A preliminary analysis conducted on a sample of 60 genes encoding ribosomal proteins in rice suggests that the association of site II motifs and *telo* boxes within their promoters is also a rule rather than an exception. Figure 2(b) illustrates the topological association of these motifs within three *Arabidopsis* promoters and the conservation of this organisation within corresponding rice promoters. For genes encoding 40S S27 proteins, the *telo* box is located downstream from the transcription initiation site as is the case for *Arabidopsis* and rice *PCNA* genes (see Figure 1a). As for the *Arabidopsis* gene encoding the translation elongation factor EF1A (Axelos *et al.*, 1989; Liboz *et al.*, 1991), the *telo* box is located upstream of the initiation site for an *Arabidopsis* and a rice 40S L4 gene. The third example (60S L27 genes) illustrates a frequent case in which a *telo* box is observed both upstream and downstream of the translation initiation site.

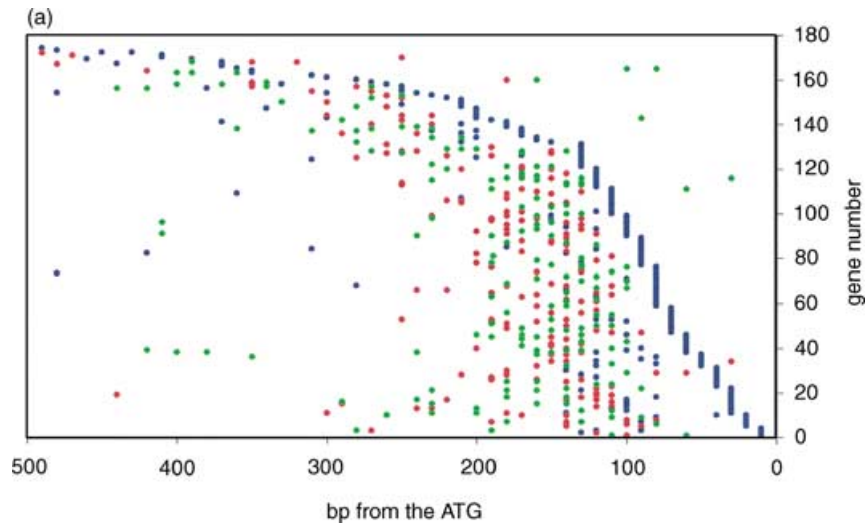
Proteins in cellular extracts and a recombinant At-TCP20 protein from the Arabidopsis TCP domain protein family bind to site II motifs

To detect proteins that interact with the site II motifs of the *At-PCNA-2* promoter, we performed an electrophoretic mobility shift analysis (EMSA) using cellular extracts prepared from exponentially growing *Arabidopsis* cell suspension cultures and a DNA fragment of the *Arabidopsis PCNA-2* promoter containing either two TGGGCC site II motifs (WT-CC) or a variant containing two TGGGCT site II motifs (WT-CT) or a mutation in these sites which was known to affect expression in primordia as probes (see Figures 1 and 3). Using the WT probes, multiple retarded complexes were formed (data not shown). To analyse these complexes in more detail, a cellular extract was fractionated by heparin sepharose filtration and each fraction was tested for its ability to bind both the WT probes and the mutated probe. These cellular extracts contained at least six distinct DNA-binding activities which, as judged by the effect of the double mutation, appear to be specific for

Figure 2. Topological organisation of *telo* boxes and site II motifs within the promoters of *Arabidopsis* genes encoding ribosomal proteins.

(a) For clarity, the 174 *Arabidopsis* ribosomal promoters harbouring one or several *telo* boxes were classified according to the location of *telo* boxes relative to the translation initiation site (blue lozenges). For each promoter, the location of TGGGCC motifs or TGGGCT motifs was indicated by red circles or green lozenges, respectively.

(b) Illustration of the topological arrangement of site II motifs and *telo* boxes and their conservation within *Arabidopsis* and rice promoters. TGGGCC are in red characters, TGGGCT in green, *telo* boxes in blue and putative TATA boxes are capitalised. The asterisks within the *Arabidopsis* promoter sequence indicate the location of the 5' end of the longest mRNA or EST *Arabidopsis* sequence of corresponding genes. For the rice L27 promoter, a putative *tef* box (Curie *et al.*, 1991) is shown (bold underlined letters).



(b)

*Ribosomal protein L27**A. thaliana* (At1g23290)

atattaggattcatataat**ggggcct**catag**ccca**atTTAATAACATAAATAACCTGTGG
 tataat**ggggcca**aatattatataaa**agccca**aaagacaatacactatctctgtttctccga
 aa**ccctag**ctgctaaatctctCTATATAAA**aacctt**taagatcatttctggttca***a**
aaccctagatctgatactaaa**ATG**

O. sativa (OSJNBa0027P10.10)

gt**ggggc**agatgctgtgcccggcgtctgataact**ggggc**tacatggcactcagagcgca
 a**agccca**tctc**agccca**gctcccactcttccattttgcccctacgctcggccacctatc
 ttaccactgacacattggccccaccaccactaccatcaattaaaccgcccgatcgcga
 tctaacggaccgcatcct**agggg**taaacagctcattctatccccacc**aaaccctaa**ac
 catctctccgcaaccgcaCTATAAAagagcttaaaaaaacctctcctctcccga**aac**
ctaactccatcttccccggcggcggcgcagcagaggatcagcc**ATG**

*Ribosomal protein S27**A. thaliana* (At5g47930)

attgctaagccgaatttctcccactcctacttgtttatgtaca**atgggctggggc**ctat
 aa**ggccca**atgttttgtgtgtcacatttctgacccacagacaTATATActtagattgaa
 gctaaagctgc*agtctccgtctcttctgctt**ctaggg**tttcatcaaccaaatcgccttc
 gccgctttccacaagcaatc**ATG**

O. sativa (P0537A05.23)

cgaattgagggagaagaacacgcgaggg**ggccca**ataacaaaaaga**agccca**taaggc
ggcccatctcccactaggggtagcaccaccgctccgatcgaatccaacggcgcaggat
 gcctcgctcagCTATAAagagctgcctctctcatttgcct**aaacccta**gccgctccac
 acccaccaccaccgctcgcctccaccgccgcccgggagaaaccacaggttaagg
 caagcgcgcccgcgctgcccgcgcccggcgaag**ATG**

*Ribosomal protein S4**A. thaliana* (At5g07090)

atTTtagctacaacaaaatataagggcgaaaaagaatgcttttagt**ggggc**ctaaatatt
 catgctgtaaaaa**ggggc**tatctaaata**agccca**ataaatttatgcgagaacaacacta
gggttttagttttcttttgaTATAAgtaaagggcgcggttaatcgcagcctc*agt
 ggaacagagcagctgcaaac**ATG**

O. sativa (P0514H03.10)

gtggctgTTTTctgTtatccaaactagTatttTgTgTcGaggaggaacGagT**ggggc**
 gaagaagatgcaaaaccagcccatttactcatcacgtgtaaaaaagaaa**agccca**aacc
ggcccaccacgga**aaacccta**acttgcgctgcgctccgcagcaTATAAcctctcccactc
 tcgctccgcatccctctctctctttggcggcggcggcgggaagagagcgcagcagac
 ggacggcgtggtggtctccccctccggcagctagctcgagctcgcaacc**ATG**

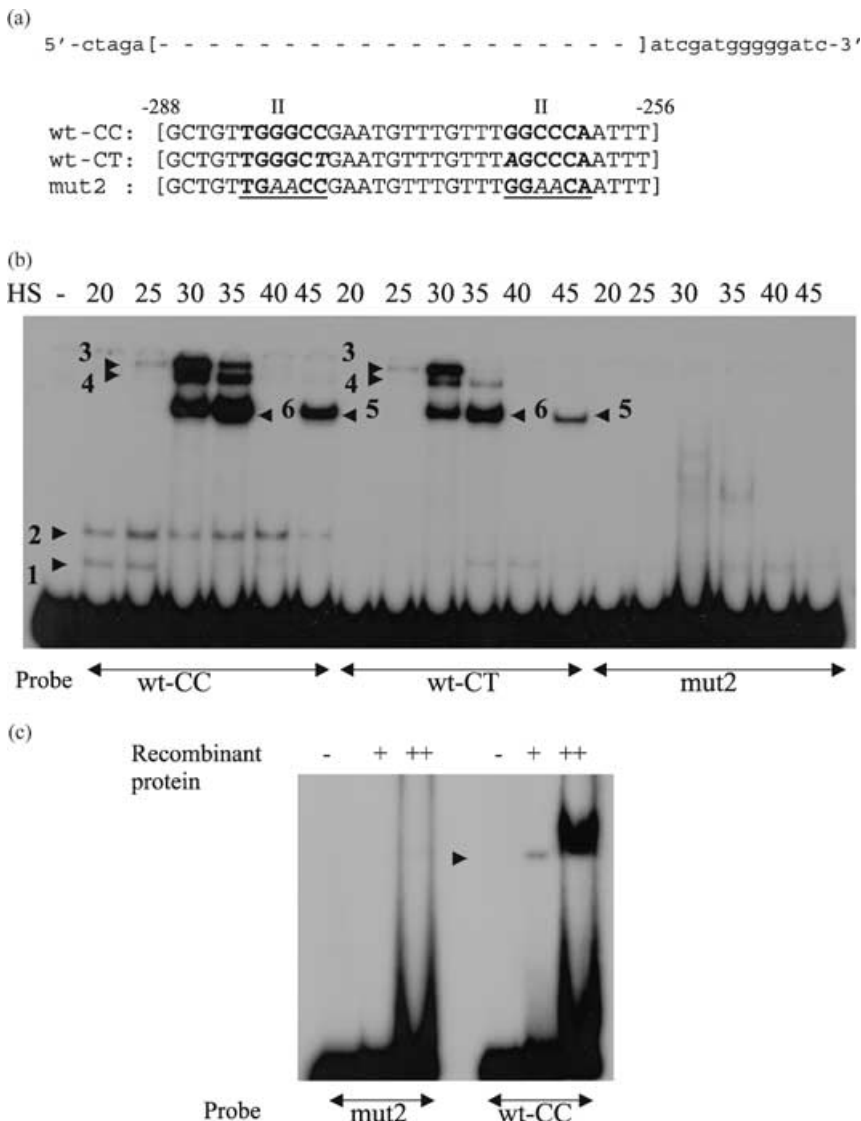


Figure 3. *At-PCNA-2* site II motif-specific protein binding. (a) The sequence of wild type and mutated probes used for electrophoretic mobility shift analysis (EMSA). Lower case letters show sequences corresponding to common regions in the three probes. For each probe the central part of the sequence is indicated in upper case letters. Motif II are in bold, and underlined in the mutant probe. (b) Binding activity of heparin sepharose fractions. Fractions are numbered above the picture. In the first lane (-) no protein was added in the binding reaction to the WT-CC probe. Specific complexes are numbered and indicated by arrowheads. The probe names are indicated below the figure. (c) MBP-TCP20 binding activity. EMSA was conducted as reported in the Experimental procedures section. The (-) lanes correspond to migration of the probes without protein; the (+) and (++) lanes correspond to migration obtained with increasing quantities of the recombinant protein. The arrow indicates the complex obtained with the WT-CC probe. The probe names are indicated below the figure.

the TGGGCC site II motif (Figure 3b). When a WT-CT probe is used, some of the activities are conserved even if the efficiency of binding seems different (arrows 3–6 in Figure 3b). However, binding of others is completely abolished (arrows 1 and 2). In rice, two proteins with a TCP domain (PCF1 and PCF2) bind to the *Ila* and *Ilb* cis-elements of the rice *PCNA* gene which are involved in meristematic tissue-specific expression of this gene (Kosugi and Ohashi, 1997). Here, we have analysed the ability of an *Arabidopsis* protein At-TCP20 (Accession no. AJ291749) related to the PCF1 and PCF2 rice transcription factors to interact with the site II motifs of the *Arabidopsis PCNA-2* promoter. The *At-TCP20* cDNA was cloned into pMAL-c2 and introduced into *Escherichia coli* strains. An EMSA was conducted with the purified recombinant fusion protein (MBP-TCP20) by using the WT-CC and the mutated probe described above.

At-TCP20 specifically interacts with the site II motifs of *PCNA-2* promoter (Figure 3c).

Physical interaction between AtPurα and At-TCP20

An *Arabidopsis* protein related to Purα in animals, AtPurα, has been recently characterised by its ability to interact with plant telomere motifs (Trémousaygue *et al.*, 1999). In animals, Purα is a DNA- and RNA-binding protein implicated in several important biological processes including the control of transcription, replication and cell cycle (for review see Gallia *et al.*, 2000). It interacts with a large number of proteins such as E2F and SP1, two transcriptional factors involved in the control of *PCNA* expression in animals. In plants, the AtPurα protein was also shown to interact with one E2F activator (Rossignol *et al.*, 2002). We thus tested the

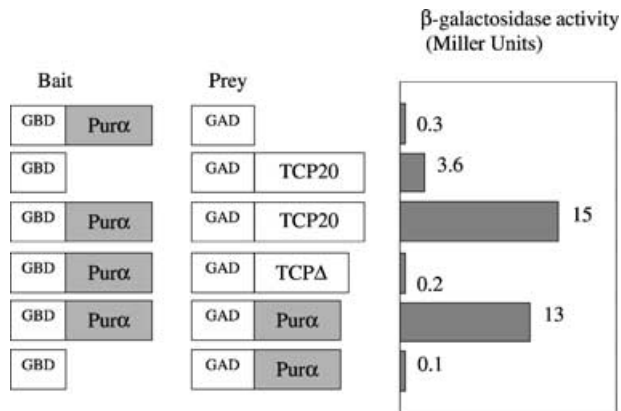


Figure 4. Physical interactions between AtPur α and At-TCP20 in yeast two-hybrid analysis. Yeast cells were co-transformed with the indicated bait and prey combinations, and a quantitative assay for β -galactosidase activity was carried out. The values presented are the average of at least six independent assays. In the GAD-TCP Δ prey construct, 184 amino acids from the C-terminal region of At-TCP20 have been deleted.

ability of AtPur α to physically interact with At-TCP20 using a two-hybrid approach. The entire coding sequence of AtPur α was fused in frame to the GAL4 DNA-binding domain in an appropriate yeast two-hybrid bait vector and the At-TCP20 coding sequence was fused in frame with an appropriate prey vector carrying the GAL4 activation domain. Co-transformed yeast cells showed a physical interaction between AtPur α and At-TCP20 as judged from both His-auxotrophy (not shown) and β -galactosidase activity (Figure 4). Moreover, the experimental controls (Figure 4) indicated that the C-terminal region of At-TCP20 was required for this interaction and that, like Pur α in animals, AtPur α is able to self-associate.

Discussion

We have characterised a ubiquitous promoter element, the site II motif (TGGGCC), required for gene expression in cycling cells in *Arabidopsis*. By searching for homologies with the core sequence of *Ila* and *Ilb* cis-acting elements of the rice *PCNA* promoter (Kosugi *et al.*, 1995), duplicated site II motifs were initially detected within the promoter of one of the two genes encoding *PCNA* in *Arabidopsis* (Figure 1). Moreover, sequence analysis of several *Arabidopsis* genes, which are upregulated during the G₁-S phase transition, such as the ribonucleotide reductase gene (ID At2g21790) or the thymidylate synthase gene (ID At4g34570), shows that site II motifs are present within their promoters (data not shown). Unexpectedly, most of the promoters of *Arabidopsis* genes encoding a ribosomal protein of the 40 or the 60S ribosomal subunit were also found to contain site II motifs (Figure 2). In this analysis, a TGGGCT variant motif was observed as frequently as TGGGCC and this led us to

propose the consensus TGGGCC/T sequence for site II motifs (Figure 2a). The same regulatory elements are found in rice ribosomal protein promoter sequences (Figure 2b). Thus, these site II motifs are probably conserved promoter control elements which act to co-ordinate the expression of *RP* genes and a set of genes upregulated at the G₁-S transition in plants. The consensus sequence we present for the site II motif differs slightly from the two PCF rice consensus sequences deduced by Kosugi and Ohashi from the random binding site selection analysis with two PCF factors, PCF2 and PCF5, except for the core sequence GGNCC (Kosugi and Ohashi, 2002). This discrepancy probably reflects the potential for different TCP proteins to act on several classes of *cis*-elements and suggests differential binding specificities for each member of the TCP protein family. The TCP proteins form a large family of 20 proteins with a conserved region predicted to form a non-canonical basic helix-loop-helix structure involved in DNA binding and dimerisation (Cubas *et al.*, 1999). Genetic and cellular approaches have shown that all the members of the TCP gene family investigated so far play a role in processes related to cell proliferation and morphogenesis (Cubas, 2002 and references therein). To our knowledge, biochemical investigations on the functions of the TCP family proteins have only been conducted on the rice PCF proteins (Kosugi and Ohashi 1997, 2002). The site II motif identified in our work specifically binds, *in vitro*, a recombinant *Arabidopsis* protein, At-TCP20 (Figure 3c), a member of this family in *Arabidopsis*. Alignment of the protein sequences from the TCP family in *Arabidopsis* confirms that the proteins form two subfamilies, and indicates that At-TCP20 belongs to a subfamily of 13 members related to rice PCF1 and PCF2 factors (Cubas, 2002; Cubas *et al.*, 1999; Hervé, unpublished). Results obtained in two-hybrid experiments indicate that At-TCP20, expressed alone as a bait in yeast two-hybrid system, is able to strongly activate the reporter gene (data not shown), whereas Kosugi and Ohashi did not detect activation properties of the PCF2 factor in the same system (Kosugi and Ohashi, 2002). Even though PCF2 and At-TCP20 seem closely related, they probably do not play the same regulatory roles, and therefore it is not surprising that they bind to different DNA sequences. From transgene expression analysis it appears that TGGGCC site II motifs are required and are sufficient for the activation of expression in cycling cells of root primordia and young leaves (Figure 1), which seems to correlate with the activation potential of the putative At-TCP20 *trans*-acting factor demonstrated in yeast. In *Arabidopsis* cellular extracts, we show that at least six DNA-binding activities specifically interact with the site II motif of the *Arabidopsis* *PCNA*-2 promoter (Figure 3b). These six DNA-binding activities might represent different members of the *Arabidopsis* TCP-PCF subfamily in homodimeric or heterodimeric configurations or even association with other partners. Indeed,

by adding deoxycholate, which is known to inhibit protein–protein interactions but not protein–DNA interactions, the six retarded complexes are disrupted (data not shown). Further work is needed to determine which of these activities contain At-TCP20 (antibodies against At-TCP20 detect proteins in all of the column fractions; data not shown). The binding pattern observed when a WT-CT probe is used instead of a WT-CC probe confirms that related proteins recognise the two probes. However, recombinant At-TCP20 proteins weakly recognise a WT-CT probe in a gel retardation assay (data not shown). The variations observed between the two profiles might be differential regulatory cues, involving both several related *cis*-acting elements and family of *trans*-acting partners.

Structural analysis reveals that the *PCNA*, *RNR*, *TS* and most of the *Arabidopsis* *RP* promoters (Figure 2) containing a site II motif also contain a telomeric motif AAACCCTA or a *telo* box. We confirm (Figure 1) the role of the *telo* boxes in the activation of expression of a set of genes induced or over-expressed in cycling cells (Manevski *et al.*, 2000; Tremousaygue *et al.*, 1999). The *telo* box is not sufficient by itself to activate gene expression but acts in synergy with several *cis*-activating sequences to potentiate their effects. A ubiquitous relationship between the control of gene expression and the telomere functions may therefore exist by analogy to that seen in yeast by RAP1, which affects gene expression, telomere structure and perhaps meiotic recombination (for reviews see Morse, 2000; Shore, 1994, 1997). In plants, such a relationship could partly explain one effect of telomere erosion on metabolism in *Arabidopsis* telomerase-deficient mutants (Fitzgerald *et al.*, 1999). Indeed, the late-generation mutants obtained by selfing plants heterozygous for a T-DNA insertion in the unique telomerase-encoding gene *AtTERT* have an extended lifespan and are unexpectedly metabolically active (Riha *et al.*, 2001). How this activation process is achieved at the molecular level remains to be established. Several genes encoding plant proteins able to bind double-stranded telomeric DNA *in vitro* have recently been characterised: *RTBP1* in rice (Yu *et al.*, 2000), *AtTBP1* and *AtPur α* in *Arabidopsis* (Hwang *et al.*, 2001, Tremousaygue *et al.*, 1999). The *Arabidopsis* *Pur α* gene which binds double-stranded or G-rich single-stranded telomeric DNA *in vitro* (Tremousaygue *et al.*, 1999) encodes a protein related to the conserved animal protein Pur α (reviewed by Gallia *et al.*, 2000). Even if these proteins have no obvious structural similarity to other described telomeric binding proteins, our knowledge about the properties of animal Pur α permits a functional parallel to be drawn with the yeast RAP1 protein. First, human Pur α has the ability to stimulate transcriptional activation by opening chromatin like the RAP1 protein (Darbinian *et al.*, 2001; Yu and Morse, 1999). The region of human Pur α which is involved in the helix-destabilising activity maps to the DNA-binding domain of this protein (Darbinian *et al.*,

2001), a region which is well conserved between animal and *Arabidopsis* Pur α (Tremousaygue *et al.*, 1999). Second, like RAP1, animal Pur α interacts with various cellular or viral transcriptional factors leading to the recruitment and/or the titration of these transactivators (for reviews see Gallia *et al.*, 2000; Morse, 2000). Notably, E2F and pRB are among these hPur α partners, and it has been shown that AtPur α also interacts with a plant E2F protein (Rossignol *et al.*, 2002). The involvement of hPur α in cell cycle control is strongly supported by transfection experiments of the protein in NIHT3-cultured cell lines at different stages of the cycle (Stacey *et al.*, 1999). Thus, AtPur α could play a role in the regulation of expression in cycling cells by interacting with *telo* box motifs. The interaction of AtPur α with the *telo* box could result in chromatin opening and in recruitment of the TCP20 activation domain close to the transcription initiation site. In this context, the conserved topological organisation of site II motifs, *telo* and TATA boxes within the *RP* promoters (Figure 2) support the notion that the *telo* box may stimulate transcription by DNA helix de-stabilisation. In addition, the physical interaction between AtPur α and At-TCP20 reported in this work supports the recruitment of At-TCP20 by AtPur α to activate gene expression.

The activation of expression driven by site II motifs in meristematic cells (Figure 1), together with the distribution of these motifs within plant ribosomal promoters (Figure 2), supports the possibility that the TCP domain transcription factors could play a key role in the over-expression of components of the translational machinery in plant cycling cells. In animals, recent evidence suggests that a number of ribosomal proteins have secondary functions independent of their involvement in protein biosynthesis. Some of these proteins function as cell proliferation regulators or as inducers of cell death (Chen and Ioannou, 1999). These results arose mainly from differential gene expression analysis and therefore suggest that transcriptional regulation is involved in these functions. In *Arabidopsis*, none of the ribosomal proteins are encoded by unique genes but rather by small multigene families (Barakat *et al.*, 2001). Analysis of site II motif and *telo* box distribution within the promoters of a same family shows that it is possible to differentiate the promoters within the family according to the distribution of these regulatory elements (data not shown). Does it mean that specialisation exists for some of the genes in processes related to cell division? Strikingly, as shown in Figure 2 for some examples, the distribution observed in *Arabidopsis* is very similar to that seen in rice. In our opinion, this characteristic emphasises the conservation of the processes involved in cell cycle regulation and provides a framework for analysing the roles of these regulatory elements. Expression studies of the *RP* genes in several environmental conditions would be useful for elucidating the precise role of site II motifs and *telo* boxes in transcriptional regulation. Such an approach should help to

unravel the biological significance of regulatory processes that control expression of both G₁/S-phase induced enzymes and ribosomal proteins in dividing cells.

Experimental procedures

Arabidopsis material

The Wassilevskija ecotype (Ws) of *Arabidopsis thaliana* was used for transformation experiments. Plants were grown in 9-h light/15-h dark period at 22°C in soil or on agar plates containing MS salts and 2% sucrose. The cell suspension culture (ecotype Colombia) used for protein extraction was grown as previously reported (Axelos *et al.*, 1992).

Bioinformatics

The *Arabidopsis* MIPS and the rice TIGR database were used to recover ribosomal gene promoter sequences and these sequences were scanned for motifs by using the 'findpatterns' GCG analysis program.

Plasmid constructions

DNA constructions were performed according to Sambrook *et al.* (1998) and were verified by sequence analysis. For plant transformation a -43 eEF1A A1 gene minimal promoter described in Tremousaygue *et al.* (1999) was used as a recipient for inserting synthetic DNA fragments, containing the *cis*-elements indicated in Figure 1. For two-hybrid experiments, the fragment corresponding to the AtPur α coding sequence was amplified from the cDNA sequence (GenBank accession no. AF136152) and introduced between the *EcoRI* and *PstI* sites of pBD-GAL4 (Stratagene ref 235612). The pAD-Pur α and pAD-TCP20 plasmids were obtained by screening an *Arabidopsis* cDNA two-hybrid library (De Veylder *et al.*, 1999). pAD-TCP20 plasmid was obtained by cloning the coding region of *TCP20* in translational fusion to the maltose-binding protein. This plasmid was deleted in its 3' end for the last 184 amino acids by exonuclease III digestion according to the manufacturer's procedure (Promega, 'Erase a base' system). The full-length cDNA of At-TCP20, from pAD-TCP20, was introduced between the *EcoRI* and *PstI* sites of pBD-GAL4 (Stratagene) to obtain pBD-TCP20. For recombinant proteins, the full-length cDNA of At-TCP20 was cloned between the *EcoRI* and *XbaI* sites of the pMalC2 vector (Biolabs) in order to obtain a fusion with the maltose-binding protein (MBP-TCP20).

Plant transformation and histochemical analysis of GUS activity

For each construct, seven independent transformants (F₁) were selected by resistance to basta and for each of them several 7-day-old F₂ plants were used for histochemical determination of GUS activity as described in Tremousaygue *et al.* (1999).

Yeast two-hybrid interaction experiments

Yeast transformation was performed according to the protocol described in the Matchmaker two-hybrid system (Clontech). Yeast Y187 cells were co-transformed with the indicated bait and prey

plasmid combinations, and then a quantitative assay for β -galactosidase activity using *O*-nitrophenyl- β -D-galactopyranoside (ONPG) as substrate was performed according to a protocol adapted from Clontech. Co-transformation for each combination was repeated three times and several independent co-transformants were tested every time.

Gel retardation experiments

Binding assays were done in a volume of 25 μ l containing 10 mM Tris-HCl pH 7.5, 1 mM DTT, 10% glycerol, 1 μ g of poly(dI-dC), 5000–10 000 cpm (1–5 ng) of ³²P-labelled probe and 5 μ l each of heparin sepharose fraction. These fractions purified from (0–40%) cell suspension protein extracts, prepared according to Manevski *et al.*, (1999), were loaded on a 5 ml column of DEAE-sepharose CL-6B (Sigma) equilibrated with buffer A (50 mM Tris-HCl, pH 8.5, 7 mM β -mercaptoethanol, 1 mM benzimidazole, 1 mM PMSF, 10% (v/v) glycerol, 50 mM NaCl). Following a wash in buffer A containing 50 mM NaCl, the fractions were eluted with a linear gradient of buffer A ranging from 50 mM NaCl to 1 M NaCl. One-millilitre fractions were collected and numbered sequentially. The recombinant TCP20 protein fused to the maltose-binding protein was purified by affinity chromatography on amylose columns (New England Biolabs). After 20 min of binding at 25°C, the free and bound DNA were separated on 6% polyacrylamide gels in 1 \times TGE (250 mM Tris-base, 1.9 M glycine, 10 mM EDTA) at a constant 150 V.

Acknowledgements

We would like to thank Dr Grimsley for critical reading of the manuscript. We thank Chantal Riou for helpful discussion about the two-hybrid system. The help of D. Douillhac for culture medium preparation is greatly appreciated.

References

- Axelos, M., Bardet, C., Liboz, T., Le Van Thai, A., Curie, C. and Lescure, B. (1989) The gene family encoding the translation elongation factor eEF1A: molecular cloning, characterization and expression. *Mol. Genet.* **219**, 106–112.
- Axelos, M., Curie, C., Mazzolini, L., Bardet, C. and Lescure, B. (1992) A protocol for transient gene expression in *Arabidopsis thaliana* protoplasts isolated from cell suspension cultures. *Plant Physiol. Biochem.* **30**, 123–128.
- Barakat, A., Szick-Miranda, K., Chang, I.F., Guyot, R., Blanc, G., Cooke, R., Delsen, M. and Bailey-Serres, J. (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol.* **127** (2), 398–415.
- Bergemann, A.D., Ma, Z-W. and Johnson, E.M. (1992) Sequence of the cDNA comprising the human *pur* gene sequence and sequence-specific single-stranded DNA-binding properties of the encoded protein. *Mol. Cell. Biol.* **12**, 5673–5682.
- Chen, F. and Ioannou, Y.A. (1999) Ribosomal protein in cell proliferation and apoptosis. *Int. Rev. Immunol.* **18**, 429–448.
- Cubas, P. (2002) Role of TCP genes in the evolution of morphological characters in angiosperms. In *Developmental Genetics and Plant Evolution* (Cronk, Q.C.B., Bateman, R.M. and Hawkins, J.A., eds), London: Taylor & Francis, pp. 247–266.
- Cubas, P., Lauter, N., Doebley, J. and Coen, E. (1999) The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J.* **18**, 215–222.

- Curie, C., Axelos, M., Bardet, C., Atanassova, R., Chaubet, N. and Lescure, B. (1993) Modular organization and developmental activity of an *Arabidopsis thaliana* eEF1a gene promoter. *Mol. Gen. Genet.* **238**, 428–436.
- Curie, C., Liboz, T., Bardet, C., Gander, E., Médale, C., Axelos, M. and Lescure, B. (1991) *Cis*- and *trans*-acting elements involved in the activation of the *Arabidopsis thaliana* A1 gene encoding the translation elongation factor eEF1a. *Nucl. Acids Res.* **19**, 1305–1310.
- Darbinian, N., Gallia, G. and Khalili, K. (2001) Helix-destabilizing properties of the human single-stranded DNA- and RNA-binding protein Pur α . *J. Cell. Biochem.* **80**, 589–595.
- De Veylder, L., De Almeida Engler, J., Bursens, J., Manevski, A., Lescure, B., Van Montagu, M., Engler, G. and Inze, D. (1999) A new D-type cyclin of *Arabidopsis thaliana* expressed during lateral root formation. *Planta*, **208**, 453–462.
- Fitzgerald, M.S., Riha, K., Gao, F., Ren, S., McKnight, T. D. and Shippen, D.E. (1999) Disruption of the telomerase catalytic subunit gene from *Arabidopsis* inactivates telomerase and leads to a slow loss of telomeric DNA. *PNAS*, **96** (26), 14813–14818.
- Gallia, G., Johnson, E. and Khalili, K. (2000) Pur α , a multifunctional single-stranded DNA- and RNA-binding protein. *Nucl. Acids Res.* **28**, 3197–3205.
- Gantt, J. and Key, J. (1985) Co-ordinate expression of ribosomal protein mRNA following auxin treatment of soybean hypocotyls. *Biochemistry*, **22**, 4131–4139.
- Gao, J., Kim, S., Chung, Y., Lee, J. and An, G. (1994) Developmental and environmental regulation of two ribosomal protein genes in tobacco. *Plant Mol. Biol.* **25**, 761–770.
- Hwang, M., Chung, I., Kang, B. and Cho, M. (2001) Sequence-specific binding property of *Arabidopsis thaliana* telomeric DNA binding protein 1 (AtTBP1). *FEBS Lett.* **503**, 35–40.
- Ito, T., Kim, G.T. and Shinozaki, K. (2000) Disruption of an *Arabidopsis* S-13 homologous gene by transposon-mediated mutagenesis causes aberrant growth and development. *Plant J.* **22** (3), 257–264.
- Kosugi, S. and Ohashi, Y. (1997) PCF1 and PCF2 specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell*, **13**, 1437–1452.
- Kosugi, S. and Ohashi, Y. (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* **30** (3), 337–348.
- Kosugi, S., Suzuka, I. and Ohashi, Y. (1995) Two of three promoter elements identified in a rice gene for proliferating cell nuclear antigen are essential for meristematic tissue-specific expression. *Plant J.* **7**, 877–886.
- Liboz, T., Bardet, C., Le Van Thai, A., Axelos, M. and Lescure, B. (1990) The four members of the gene family encoding the *Arabidopsis thaliana* translation elongation factor eEF1a are actively transcribed. *Plant Mol. Biol.* **14**, 107–110.
- Manevski, A., Bardet, C., Trémoussaygue, D. and Lescure, B. (1999) Characterization and properties of heteromeric plant protein complexes that interact with the *tef* *cis*-acting elements in both RNA polymerase II-dependent promoters and rDNA spacer sequences. *Mol. Gen. Genet.* **261**, 892–900.
- Manevski, A., Bertoni, G., Bardet, C., Trémoussaygue, D. and Lescure, B. (2000) In synergy with various *cis*-acting elements, plant interstitial telomere motifs regulate gene expression in *Arabidopsis* root meristems. *FEBS Lett.* **483**, 43–46.
- Meijer, M. and Murray, J. (2001) Cell cycle controls and the development of plant form. *Curr. Opin. Plant Biol.* **4**, 44–49.
- Morse, R.H. (2000) RAP, RAP, open up New wrinkles for RAP1 in yeast. *Trends Genet.* **16** (2), 51–53.
- Ohnishi, N., Kodama, H., Ando, S. and Komamine, A. (1990) Synthesis of protein and mRNA is necessary for the transition of suspension-cultured *Catharanthus roseus* cells from G₁ to the S phase of the cell cycle. *Physiol. Plant.* **80**, 95–101.
- Pennacchio, L. and Rubin, E. (2001) Genomic strategies to identify mammalian regulatory sequences. *Nat. Genet.* **2**, 100–109.
- Regad, F., Hervé, C., Marinx, O., Bergounioux, C., Trémoussaygue, D. and Lescure, B. (1995) The *tef1* box, a ubiquitous *cis*-acting element involved in the activation of plant genes that are highly expressed in cycling cells. *Mol. Gen. Genet.* **248**, 703–711.
- Regad, F., Lebas, M. and Lescure, B. (1994) Interstitial telomere repeats within the *Arabidopsis thaliana* genome. *J. Mol. Biol.* **239**, 163–169.
- Richards, E. and Ausubel, F. (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell*, **53**, 127–136.
- Riha, K., McKnight, T., Griffing, L. and Shippen, D. (2001) Living with genome instability: plant responses to telomere dysfunction. *Science*, **291**, 1797–1800.
- Riou-Khamlichi, C., Menges, M., Healy, S.M.J. and Murray, J.A.H. (2000) Sugar control of plant cell cycle: differential regulation of D-type cyclin gene expression. *Mol. Cell. Biol.* **20** (13), 4513–4521.
- Rossignol, P., Stevens, R., Perennes, C., Jasinski, S., Cella, R., Trémoussaygue, D. and Bergounioux, C. (2002) AtE2F-a and AtDP-a, members of the E2F family of transcription factors, induce *Arabidopsis* leaf cells to re-enter S phase. *Mol. Genet. Genomics*, **266** (6), 995–1003.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Scheer, I., Ludevid, M., Regad, F., Lescure, B. and Pont-Lezica, R. (1997) Expression of a gene encoding a ribosomal p40 protein and identification of an active promoter site. *Plant Mol. Biol.* **35**, 905–913.
- Shore, D. (1994) RAP1: a protein regulator in yeast. *TIG*, **10**, 408–412.
- Shore, D. (1997) Telomere length regulation: getting the measure of chromosome ends. *Biol. Chem.* **378** (7), 591–597.
- Stacey, D.W., Hitomi, M., Kanovsky, M., Gan, L. and Johnson, E.M. (1999) Cell cycle arrest and morphological alterations following microinjection of NIH3T3 cells with Pur alpha. *Oncogene*, **29**, 4254–4261.
- Trémoussaygue, D., Manevski, A., Bardet, C., Lescure, N. and Lescure, B. (1999) Plant interstitial telomere motifs participate in the control of gene expression in root meristems. *Plant J.* **20**, 553–561.
- Van Lijsebettens, M., Vanderhaeghen, R., De Block, M., Bauw, G., Villaroel, R. and Van Montagu, M. (1994) An S18 ribosomal protein gene copy at the *Arabidopsis* PFL locus affects plant development by its specific expression in meristems. *EMBO J.* **13**, 3378–3388.
- Weijers, D., Franke-van Dijk, M., Venken, R.-J., Quint, A., Hooykaas, P. and Offringa, R. (2001) An *Arabidopsis* minute-like phenotype caused by a semi-dominant mutation in a ribosomal protein S5 gene. *Development*, **128**, 4289–4299.
- Yu, E.Y. and Morse, R.H. (1999) Chromatin opening and transactivator potentiation by RAP1 in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **19** (18), 5279–5288.
- Yu, E.Y., Kim, S.E., Kim, J.H., Ko, J.H., Cho, M.H. and Chung I.K. (2000) Sequence-specific binding DNA recognition by the Myb-like domain of plant telomeric protein RTTBP1. *J. Biol. Chem.* **275** (31), 24208–24214.