

Functional architecture of a late pollen promoter: pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements

N. Bate¹ and D. Twell*

Department of Biology, University of Leicester, University Road, Leicester LE1 7RH, UK (*author for correspondence); ¹Present address: Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK

Received 1 July 1997; accepted in revised form 13 February 1998

Key words: CaMV 35S, *Nicotiana tabacum*, pollen, reporter genes, transcription

Abstract

The tomato *lat52* gene encodes an essential cysteine-rich protein preferentially transcribed in the vegetative cell during pollen maturation. Detailed analyses of the identity, organization and role of *cis*-regulatory elements in controlling the precise developmental and tissue-specific expression of *lat52* during pollen development were performed. Analysis of a series of 5' promoter deletion mutants stably introduced into tobacco demonstrated differential developmental activation of deletion mutants during pollen development. All major *cis*-regulatory elements required for pollen-specific transcription were located within the upstream region –492 to –52. This region was shown to comprise three independent activator domains A, B and C, each sufficient to activate the minimal CaMV 35S promoter in a pollen-specific manner. 5' deletion and gain of function approaches were used to show that domain A and the previously defined motif PBII (sub-domain B1) were largely redundant in the presence of downstream sequences in mature pollen. Within domain B two novel pollen-specific sub-domains B2 and B3 were identified. Within domain C, the activity of the PBI motif (sub-domain C1) was shown to be strictly dependent upon a downstream 20 bp pollen-specific activator unit –72 to –52 (sub-domain C2), containing two novel co-dependent regulatory elements AGAAA and TCCACCATA. These results demonstrate that transcriptional activation of *lat52* is controlled by a complex of pollen-specific *cis*-regulatory elements which cooperate to achieve maximum levels of gene expression throughout pollen maturation. Alternative models of the interaction of identified *cis*-regulatory elements with putative *trans*-acting factors within the *lat52* promoter and their developmental utilization are presented.

Introduction

The pathway of pollen maturation provides an accessible and well characterised system in which to study the tissue-specific and developmental control of gene expression. At maturity the pollen grain contains about 20 000 unique transcripts of which 2000 are estimated to be specifically or preferentially expressed in pollen [25, 39, 40]. The similar expression patterns of a number of these 'late pollen genes' suggests a common mechanism of regulation since transcripts appear at pollen mitosis I and accumulate in the maturing pollen grain up to anthesis [30]. In order to further dissect the pathway of late pollen gene regulation it

is important to functionally define the *cis*-regulatory elements responsible for coordinate and specific activation in the developing pollen grain. One of the most well understood late pollen genes in terms of its regulatory properties is the late anther tomato, *lat52* gene [18], which shows significant homology to cDNAs isolated from maize [12], rice [42], sorghum [22], lilac [1] and olive [17]. The *lat52* gene encodes an abundant cysteine-rich protein which is distantly related to Kunitz trypsin inhibitor family [19]. The *lat52* homologue in olive encodes the major pollen allergen, OLE EI, which is localised to the endoplasmic reticulum [24] and accounts for ca. 1% of total dry mass of mature pollen [38]. An essential gametophytic role for

the LAT52 protein in pollen hydration and tube growth has been demonstrated by antisense-mediated down-regulation in tomato [20].

Previous studies of native *lat52* mRNA expression in tomato demonstrated an essentially pollen-specific expression pattern with only low levels of transcripts detectable in the anther wall [37] and in petals [18]. During gametophyte development *lat52* transcripts were first detectable in spores undergoing pollen mitosis I, leading to a substantial increase in mature pollen [18]. The specificity of transcriptional activation was further demonstrated since the *lat52* 5'-flanking region (−492 to +110) was shown to direct pollen-specific GUS activity in tomato, tobacco and *Arabidopsis* [35] and pollen-specific cell ablation when linked to the cytotoxic diphtheria toxin A chain [31]. Furthermore, the developmental accumulation of GUS activity in pollen of transgenic plants closely reflected the accumulation of the native *lat52* transcript [33, 6].

Sequence analysis of the *lat52* promoter, led to the identification of the reiterated PB core motif TGTG-GTT, (PBI, II and III) which is closely related to the *rbcS*-3A Box II *cis*-regulatory element [9, 16]. Analysis of *lat52* 5' promoter deletion mutants in tomato led to the identification of two pollen-specific activator regions: −492 to −125 (containing PBIII and PBII) and −124 to −72 (containing PBI). This analysis also defined the minimal proximal promoter sufficient to direct pollen-specific expression as the region −71 to +110. Further gain of function analyses demonstrated that both PBII (−194 to −176) [36] and the 30 bp proximal promoter region (−84 to −55) [7] were sufficient to direct pollen-specific expression.

Here the results of experiments designed to further elucidate the precise sequences and organization of *cis*-regulatory elements required for the developmental and cell-specific regulation of the *lat52* gene are described. Analyses of promoter deletion mutants demonstrated a differential utilization of regulatory domains during pollen development. Three independent pollen-specific activator domains and functionally redundant sequences were defined. Minimal sequences sufficient for pollen-specific expression were localised to a novel bipartite 20 bp sequence. The results are discussed in relation to working models of the interaction of regulatory modules involved in the control of gene expression during pollen development.

Materials and methods

Plasmid constructions

Construction of plasmids containing the *lat52* promoter 5' deletion mutants (−492 to +110, −100 to +110, −71 to +110 and −41 to +110) fused to the *Escherichia coli* β -glucuronidase gene *uidA* (*gus*) and the nopaline synthase 3' region (*nos3'*) have been described previously [36]. The *lat52-gus-nos3'* fusions were excised as *SalI/EcoRI* fragments and cloned into the binary vector pBIN19 [3]. *lat52* promoter 5' deletion mutants (−492, −225, −145, −100, −71, −41 and −17) fused to the luciferase coding region (*luc*) [21] and the CaMV 35S 3' terminator region (*C3'*) [27] were created by subcloning *lat52* promoter regions from *lat52-gus-nos3'* cassettes [36] as *SalI/NcoI* fragments into a promoterless *luc-c3'* cassette in pBluescript KS⁺ (Stratagene). The *lat52* promoter regions −194 to +110 and −176 to +110 were amplified from the template pNBL52-5 [2] using the polymerase chain reaction (PCR) in conjunction with specific 5' oligonucleotides containing a *KpnI* site at the relevant 5' border sequence of the *lat52* promoter and a 3' oligonucleotide complementary to the non-sense strand of *luc*. The generated PCR products were digested with *KpnI/NcoI* and cloned into the above *luc-c3'* cassette. The *lat52* promoter 5' deletion +1 to +110 was created by subcloning the *lat52* 5'-*utr-luc-c3'* cassette from pNBL52-27 [2] as a *XhoI/HindIII* fragment into pBluescript KS⁺.

Plasmids used for the substitution mutational analysis of the region −100 to −42 of the *lat52* promoter were created as follows. The 2 bp series of substitution mutants throughout PBI were introduced using PCR and degenerate oligonucleotides. The amplified region −100 to +110 containing the introduced 2 bp mutation was cloned as a *KpnI/NcoI* fragment into the *luc-c3'* cassette. The block substitution mutants of the region −84 to −42 were created by amplifying the promoter region downstream of the desired mutation using PCR. This fragment was then cloned into a *luc-c3'* cassette as a *SalI/NcoI* fragment. The upstream region of the desired mutation was then amplified using PCR and cloned into the *SalI* site.

Hybrid CaMV 35S-*lat52* mutants were constructed by amplifying the duplicated CaMV 35S region (−418 to −90/−418 to −78) from the plasmid pRTL2LUC [29] using PCR. The duplicated CaMV 35S enhancer region was cloned directly upstream of the appropriate *lat52* 5' promoter deletion mutant-*luc-c3'* cas-

sette as a *KpnI/SalI* fragment. For the gain of function analysis the minimal CaMV 35S-TEV-LUC-C3' cassettes were first created. The region containing –45 to +1 CaMV 35S-TEV 5'-UTR was PCR-amplified from pRTL2LUC and cloned into a promoterless *luc-c3'* cassette as a *KpnI/NcoI* fragment to give the plasmid pNBC-45 LUC. The region containing –90 to +1 CaMV 35S-TEV 5'-UTR was amplified from pRTL2LUC using PCR and cloned into pNBC-45LUC as a *Sall/NcoI* fragment to give the plasmid pNBC-90 LUC. Hybrid *lat52*-CaMV 35S promoter fusions were then constructed by synthesising fragments of the *lat52* promoter as before, these were cloned as *KpnI/SalI* fragments directly upstream of –90 CaMV 35S-*luc* or –45 CaMV35S-*luc*.

All PCR fragments were entirely sequenced and no mutations were found except for those intentionally introduced within primer sequences.

Microprojectile bombardment

Equal amounts of supercoiled test plasmid with the appropriate reference plasmid were co-precipitated onto tungsten particles and bombarded into *Nicotiana tabacum* cv. Samsun pollen or leaves essentially as described by Twell *et al.* [34]. Pollen was incubated for 16 h on solid germination medium [28] prior to protein extraction. The reference plasmids used for pollen and leaves were pLAT52-7 (*lat52-gus-nos3'*) [33] and pRTL2GUS (CaMV 35S5'-TEV-*gus*-CaMV 35S3') [5] respectively.

Plant transformation

Binary vectors were transferred directly into *Agrobacterium tumefaciens* strain LBA4404 (pAL4404) [35]. Tobacco leaf discs were transformed with *A. tumefaciens* as previously described [6]. Shoots were rooted on solid MS30 medium containing 50 mg/l kanamycin, transferred to soil and grown to maturity under greenhouse conditions.

Transient reporter gene expression assays

These were performed as previously described [36]. Luciferase activity was measured as total light emission in 10 s after addition of D-luciferin using a Berthold Clinilumat luminometer. GUS activity was measured as the change in fluorescence of the reaction product, 4-methylumbelliferone (4-MU), per hour; this was measured at an excitation of 365 nm and an emis-

sion of 455 nm using the Perkin Elmer luminescence spectrometer LS50. Test promoter activity was derived from a ratio of LUC to GUS activity and normalised to a control plasmid. Two independent experiments were performed representing six independent bombardments for each plasmid.

Histochemical localisation of GUS activity

Tissue samples were incubated at room temperature in the presence of 1 mM X-gluc and 0.5 mM potassium ferricyanide for 2 to 16 h.

Fluorometric quantification of GUS activity

GUS activity was measured as described above except plant tissue samples were ground using a pestle and mortar in the presence of GUS extraction buffer for 1 min. The specific activity of GUS in each sample was then calculated as pmol 4-MU formed per minute per mg total protein. Protein concentration in plant extracts was determined using Bradford reagent (BioRad) according to the manufacturer's instructions.

Results

The minimal promoter proximal unit sufficient for pollen-specific expression of lat52 resides within the region –41 to +110

Previous analysis defined the minimal promoter proximal unit sufficient for pollen-specific gene expression in transgenic tomato as the region –71 to +110 [36]. Using the biolistic transient expression system this was further delimited to the region –41 to +110 [36]. To confirm the significance of these findings in transgenic plants and further define upstream regulatory domains a series of 5' promoter deletion mutants linked to *gus* were introduced into tobacco. Twelve to fourteen independent transgenic lines were generated for each construct and GUS activity was determined in extracts prepared from mature pollen, leaves and T₂ seedlings. GUS activity was also determined in extracts prepared from ten independent CaMV 35S-*gus* transformants and a similar number of untransformed plants (Figure 1A).

In pollen, deletion of the *lat52* promoter from –492 to –100 decreased GUS activity 14-fold, deletion from –100 to –71 decreased activity a further 15-fold while removal of 30 bp from –71 to –41 led to a further 3-

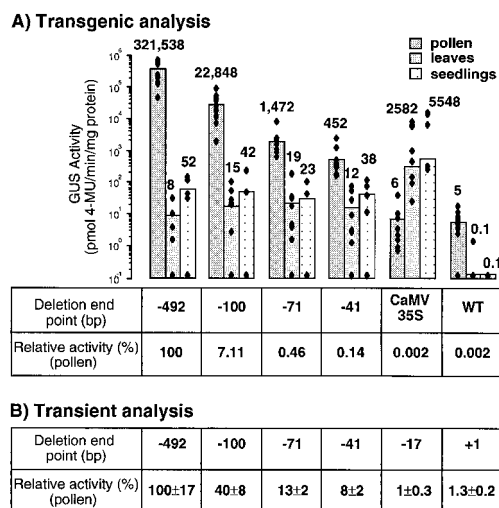


Figure 1. A. Analysis of *lat52* 5' promoter deletion mutants in transgenic tobacco plants. The scatter histograms show GUS activities determined in pollen, leaves and T₂ seedlings of transgenic plants harbouring either *lat52* 5' promoter deletion mutant-*gus* or CaMV 35S-*gus* fusions. Also shown are scatter histograms for GUS activities determined in pollen, leaves and T₂ seedlings of non-transformed plants. Each symbol represents the value derived from the analysis of a single independent transgenic plant, with the mean values indicated by a shaded box. The data summarized below represents the relative activity (%) in pollen determined from the mean GUS activities for each *lat52* 5' promoter deletion mutant-*gus* or CaMV 35S-*gus* fusion and for untransformed plants. B. Transient expression analysis of *lat52* 5' promoter deletion mutants in pollen. The activities of test plasmids containing *lat52-luc-C3'* fusions are shown relative to the activity of the largest 5' deletion mutant (–492 to +110).

fold drop in activity, which was still ca. 75-fold above that of the CaMV 35S-*gus* and untransformed controls. In leaves and seedlings low levels of GUS activity were detectable, but histochemical analysis did not result in visible staining, suggesting that low level sporophytic expression was dispersed rather than localised to a few specific cell types. In plants harbouring the shortest (–41) *lat52* deletion, deriving the ratio of mean GUS activity in pollen versus seedlings or leaves demonstrated preferential expression in pollen by 12- and 38-fold respectively. Therefore, the shortest promoter proximal region sufficient to direct preferential expression in pollen was defined as the region –41 to +110.

To independently confirm the function of the upstream activator regions and delimit functional sequences within the minimal promoter region, 5' deletion mutants were linked to the luciferase (*luc*) gene [21] and transiently expressed in pollen (Figure 1B). Deletion from –492 to –100 decreased LUC activity by 2.5-fold, deletion from –100 to –71 decreased

activity a further 3-fold and deletion from –71 to –41 decreased activity a further 1.5-fold. Further deletion to –17, removing the first putative TATA box led to an 8-fold decrease in activity, but deletion to +1, which removed a second putative TATA box did not lead to a further decrease in activity. These data confirm the function of three upstream activator regions and demonstrate that the TATA proximal region –41 to –18 is essential for expression in pollen.

On the strength of these data the *lat52* promoter was divided into four functional units: a major activator region –492 to –101 (93% of total activity), a minor activator region –100 to –72 (6.7% of total activity) and two minimal activator regions –71 to –42 (0.3% of total activity) and –41 to +110 (0.14% of total activity).

Ectopic activation of lat52 promoter deletion mutants by a heterologous enhancer

To investigate whether the observed low level of sporophytic expression (Figure 1A) was the result of genomic position effects or to weak constitutive *lat52* promoter activity, two sets of constructs were transiently expressed in leaves. The first consisted of *lat52* 5' promoter deletion mutants –100, –71, –41 and –17, whilst the second consisted of a duplicated CaMV 35S enhancer region inserted upstream of each *lat52* 5' promoter deletion.

The –100, –71 and –41 promoter deletions enhanced LUC activity 4- to 5-fold above the TATA-less –17 control (Figure 2). Addition of the CaMV 35S enhancer stimulated LUC activity 20–33-fold, but the relative differences between each deletion remained approximately the same. These results confirm weak transcriptional activation in leaves, and suggest that no strong sporophytic repressor elements reside within the region –100 to +110. These results further suggest that TATA proximal sequences in the region –41 to –18 do not preferentially function in pollen, since deletion of this region reduced LUC activity to a similar degree in both leaves and pollen (Figure 1B).

Differential utilization of lat52 activator regions during pollen development

To investigate the developmental utilization of each upstream activator region the accumulation of GUS activity in developing pollen was analyzed in three transgenic lines harbouring each *lat52* promoter deletion mutant. Since negligible levels of GUS activity

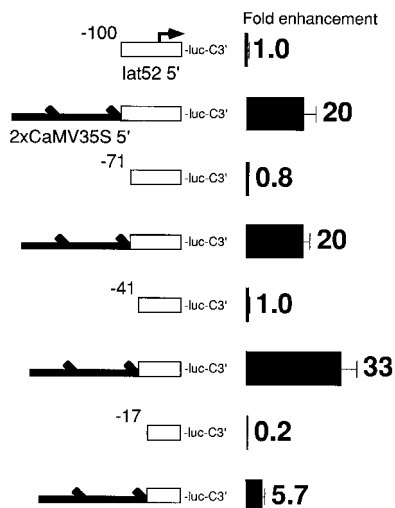


Figure 2. Transient expression analysis of *lat52* 5' promoter deletion mutants in leaves. Relative activities of test plasmids were normalised to the activity of the *lat52* 5' deletion mutant, -100 to +110.

were detectable in the anther wall [35], GUS activity was determined in extracts of anthers at precise developmental stages defined by bud length [32].

Plants harbouring -492 *lat52-gus* showed a sharp increase in GUS activity in anthers from ca. 16 mm buds (pollen-mitosis I), with a continuous rapid accumulation during pollen maturation (Figure 3). Plants harbouring -100 *lat52-gus* showed a sharp increase in GUS activity only at a later stage in pollen development (bud length ca. 30 mm) with a similar rapid accumulation during maturation. In contrast, plants harbouring -71 *lat52-gus* showed a steady increase in activity throughout pollen development. No differences were detected in the timing of GUS activation between plants harbouring the -71 or -41 *lat52-gus* fusions (data not shown). These analyses demonstrate that the *lat52* promoter is regulated by at least two stage-specific activator regions -492 to -101 and -100 to -72 during pollen development.

The three upstream domains of the lat52 promoter are each sufficient to activate a heterologous core promoter in a pollen-specific manner

The results presented in Figures 1 and 2 strongly suggest that all sequences contributing to pollen-specific transcriptional control reside within the region -492 to -42. Therefore, this region was sub-divided into three domains: domain A (-492 to -226) domain B (-225 to -101) and domain C (-100 to -52) (Fig-

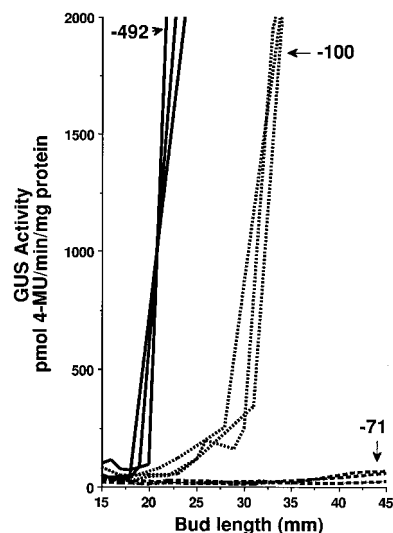


Figure 3. Developmental regulation of the -492, -100 and -71 *lat52* 5' promoter deletion mutants in transgenic tobacco. The GUS activities in extracts prepared from anthers isolated from buds of different length are shown for: three individual lines harbouring the *lat52* 5' promoter deletion -492 to +110-*gus-nos3'* fusion (black lines), three individual lines harbouring the *lat52* 5' promoter deletion -100 to +110-*gus-nos3'* fusion (short dashed lines) and three individual lines harbouring the *lat52* 5' promoter deletion -71 to +110-*gus-nos3'* (long dashed lines).

ure 4). To investigate their individual and combinatorial activities single, mutually exclusive domain pairs and all three linked domains were cloned upstream of the -90 CaMV 35S promoter and transiently expressed in pollen and leaves (Figure 5). In leaves domains A, B or C individually or in combination did not enhance -90 CaMV 35S. However, in pollen domains A, B and C were each sufficient to activate -90 CaMV 35S by 6-fold, 4-fold and 11-fold respectively. Enhancement by domains A+B was approximately additive (8-fold), whereas domains B+C synergistically enhanced activity to a level (26-fold) similar to that of domains A+B+C together (24-fold). These analyses demonstrate that domains A, B or C are each sufficient to enhance transcription in a pollen-specific manner, and suggest that cooperative interactions occur between domains B and C and that the activity of domain A is redundant in the presence of domains B+C in mature pollen.

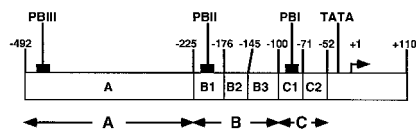


Figure 4. Functional map of the *lat52* promoter defined by transgenic and transient expression analysis.

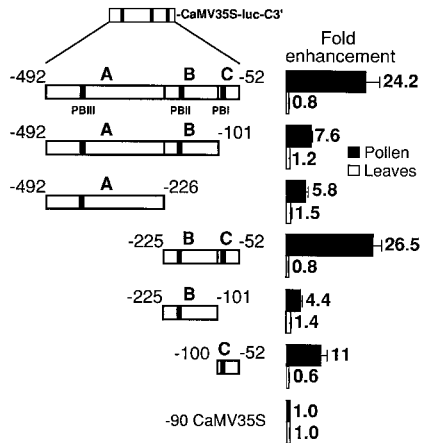


Figure 5. Activation of the -90 CaMV 35S core promoter by *lat52* promoter fragments. The relative activities of test plasmids containing chimaeric *lat52*/ -90 CaMV 35S-*luc-C3'* gene fusions were determined in pollen and leaves in the transient expression assay. The fold enhancement of each test plasmid was calculated from the ratio of LUC/GUS normalised to that of the -90 CaMV 35S-*luc-C3'* control.

Dissection of domain B reveals two pollen-specific activator sub-domains, B2 and B3

A gain of function assay was used to determine the location of regulatory elements within domain B. Domain B promoter fragments were fused to the -90 CaMV 35S promoter and transiently expressed in pollen and leaves. Domain B enhanced -90 CaMV 35S activity about 4-fold in a pollen-specific manner. Progressive 5' deletions from -225 to -176 (deleting PBII) and targeted mutation of PBII did not significantly affect pollen-specific enhancement, but further deletion to -145 reduced enhancement to 2.5-fold (Figure 6A). These data define two new novel pollen-specific sub-domains, B2 (-176 to -146) and B3 (-145 to -101) and show that PBII is not utilized in the presence of these downstream sequences. To independently confirm an active role for these sequences each region was progressively removed from the native promoter. While deletion of domain A, or domain B1 did not significantly reduce promoter activity, deletion of domains B2 and B3 progressively decreased

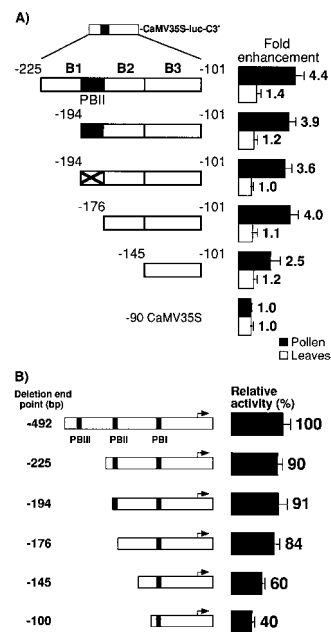


Figure 6. A. Targeted mutagenesis of the *cis*-regulatory element PBII and 5' deletion analysis of the *lat52* promoter domain B. The relative activities of test plasmids containing chimaeric *lat52*/ -90 CaMV 35S-*luc-C3'* gene fusions were determined in pollen and leaves in the transient expression assay. The fold enhancement of each test plasmid was calculated from the ratio of LUC/GUS normalised to that of the -90 CaMV 35S-*luc-C3'* control. B. Transient expression analysis of *lat52* 5' promoter deletion mutants. Relative activities of test plasmids containing *lat52*-*luc-C3'* fusions were calculated from a ratio of LUC:GUS normalised to the activity of the largest 5' deletion mutant (-492 to $+110$).

promoter activity (Figure 6B). This confirms an active role for the sub-domains B2 and B3 and further highlights the functional redundancy of domains A and B1 when linked to downstream sequences.

Mutational analysis of the cis-regulatory element PBI

Domain C contains the sequence motif PBI (TGTTGTT) flanked by AT-rich sequences, TTATTGTGGTTAT [33] such that the 3' half differs by only 1 bp from the box II motif (GGTTAA) of the light responsive *rbcS-3A* gene [9]. To investigate whether the same sequence motifs operate as in box II, a series of 2 bp scanning mutations through the PBI region were transiently expressed in pollen (Figure 7A). Mutation of the central GG residues of PBI reduced expression ca. 10-fold, while mutation of the 5'-adjacent GT residues resulted in a 2-fold reduction. In contrast, mutation of the flanking pairs of nucleotides resulted in a 2–3-fold stimulation in activity. Thus, PBI consists of a GTGG

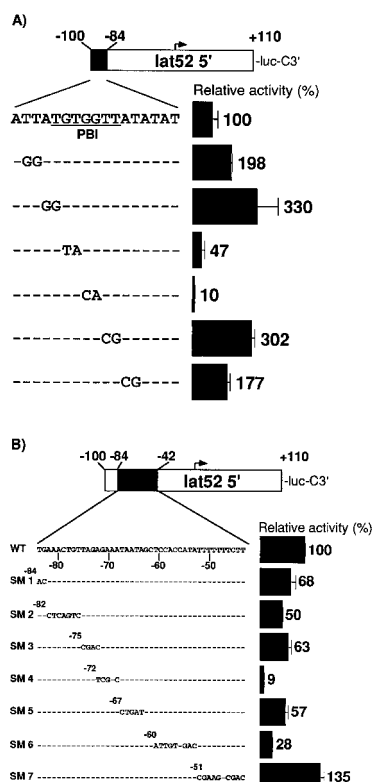


Figure 7. A. A 2 bp linker scan analysis of the pollen-specific cis-regulatory element PBI. Test plasmids containing single 2 bp linker scan mutations throughout the region -99 to -88 of the -100 to $+110$ *lat52* 5' promoter deletion-*luc-C3'* gene fusion were co-bombarded into pollen. Relative activities of test plasmids containing mutant *lat52-luc-C3'* fusions were calculated from a ratio of LUC:GUS normalised to the activity of the wild type -100 to $+100$ *lat52* 5' promoter deletion-*luc-C3'* gene fusion. B. Targeted substitution mutagenesis of domain C of the *lat52* promoter. Test plasmids containing single block substitution mutations throughout the region -84 to -42 of the -100 to $+110$ *lat52* 5' promoter deletion-*luc-C3'* gene fusion were co-bombarded into pollen. Relative activities of test plasmids containing mutant *lat52-luc-C3'* fusions were calculated from a ratio of LUC:GUS normalised to the activity of the wild-type -100 to $+110$ *lat52* 5' promoter deletion-*luc-C3'* gene fusion.

core motif, required for wild-type levels of activity, flanked on either side by TTAT motifs which negatively affect expression. Given that box II mutations which prevent GT-1 binding *in vitro* are also likely to reduce expression *in vivo* [8, 23], these results highlight a different functional core sequence for PBI. The only functional similarity is the negative effect of mutating the two central GG residues; mutations either side gave different results for both elements.

The region -84 to -41 contains two novel cis-regulatory sequences

To further define pollen-specific regulatory elements within domain C substitution mutations (SM1 to 7) were introduced throughout the region -84 to -42 and their activities assayed in pollen (Figure 7B). Substitution mutants SM1, 2, 3, and 5 decreased expression by 2-fold or less compared with the wild-type (-100) promoter, whereas SM7 slightly increased expression. The most dramatic effects were observed for SM4 and SM6 which decreased promoter activity by 10- and 4-fold respectively. These analyses define the location of two novel positive regulatory sequences within domain C, AGAAA (-72 to -68) and TCCACCATA (-60 to -52), both of which are flanked by sequences which positively influence promoter activity albeit to a lesser extent.

Domain C activates the minimal CaMV 35S promoter independent of distance

The activating properties of domain C were further investigated by placing it upstream of -45 and -90 CaMV 35S, allowing it to be tested in its original position with respect to the TATA box and at 4.5 turns of the DNA helix further upstream respectively. To confirm the function of PBI in this context two otherwise identical constructs were tested in which PBI was mutated (GTGG to TACA; Figure 7A). Domain C similarly activated both -45 and -90 CaMV 35S in a pollen-specific manner (Figure 8A), demonstrating that it is similarly functional when moved 45 bp upstream. Since mutation of PBI in this context reduced, but did not abolish activation, these results further suggest independent function of downstream sequences within domain C (Figure 7B).

The region -71 to -52 within domain C defines a minimal pollen-specific functional unit

To define the minimal independent functional units within domain C, mutant derivatives were fused to the -45 CaMV35S promoter in a gain of function assay (Figure 8B). Domain C enhanced activity by 11.5-fold in a pollen-specific manner. Although 5' deletion of PBI from domain C decreased enhancement, a fragment containing PBI (-100 to -72 ; sub-domain C1) failed to activate -45 or -90 CaMV 35S (Figure 8B and data not shown). Deletion of domain C to -71 had no further effect, but deletion to -61 (removing

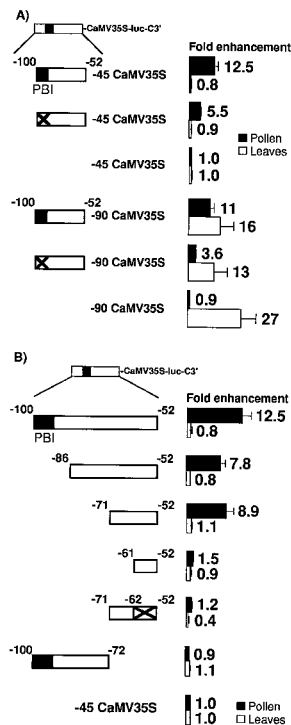


Figure 8. A. Pollen-specific activation of the -45 and -90 CaMV 35S core promoters by *lat52* promoter domain C. *lat52* promoter fragments containing domain C (-100 to -52) or domain C with the *cis*-regulatory element PBI mutated (black cross) were fused upstream of the -45 and -90 CaMV 35S minimal promoters. The relative activities of test plasmids containing chimaeric *lat52*/ -45 - or *lat52*/ -90 CaMV 35S-*luc*-C3' gene fusions were determined in pollen and leaves in the transient expression assay. The fold enhancement of each test plasmid was calculated from the ratio of LUC/GUS normalised to that of the -45 CaMV 35S-*luc*-C3' control. B. Activation of the -45 CaMV 35S core promoter by proximal *lat52* promoter fragments. The relative activities of test plasmids containing chimaeric *lat52*/ -45 CaMV 35S-*luc*-C3' gene fusions were determined in pollen and leaves in the transient expression assay. The fold enhancement of each test plasmid was calculated from the ratio of LUC/GUS normalised to that of the -45 CaMV 35S-*luc*-C3' control.

AGAAA) abolished enhancement. Targeted mutation of TCCACCATA within the fragment -71 to -52 also abolished enhancement. These data demonstrate that PBI (sub-domain C1) function is dependent on the presence of the GAAA and/or TCCACCATA motifs (sub-domain C2), and that these motifs function in a co-dependent manner. Thus, the minimal functional pollen-specific unit within domain C is defined as the 20 bp sequence GAAATAATAGCTCCACCATA (-71 to -52 ; sub-domain C2), containing the underlined bipartite regulatory sequences.

Discussion

This analysis of the identity, organization and role of *cis*-regulatory sequences within the *lat52* promoter provides new insight into the modular organization and developmental regulation of a pollen-specific transcription unit.

Pollen-specific transcriptional regulatory elements reside upstream of -41

In transgenic plants upstream domains A+B+C (-492 to -42) together accounted for 99.9% of the activity of the *lat52* promoter in pollen. However, the TATA proximal domain -41 to $+110$, which includes the entire 5'-UTR ($+1$ to $+110$), was still preferentially expressed in pollen by 10–40-fold compared with vegetative tissues. Two lines of evidence suggest that this preferential expression is largely accounted for by a translational mechanism. First, the *lat52* 5'-UTR has previously been shown to contain a strong pollen-specific translational enhancer which increases translational yield ca. 13-fold [2], and second, deletion of TATA proximal sequences (-41 to -18) led to a similar decrease in promoter activity in both pollen and leaves, suggesting that these sequences are not sufficient to support differential transcription. Thus, pollen-specific transcriptional regulatory elements are located exclusively within the upstream region -492 to -42 .

Differential utilization and functional redundancy of lat52 upstream activator domains during pollen development

Transgenic analysis demonstrated that during early pollen development domains A+B together strongly enhanced promoter activity, whereas domain C was only minimally active. In contrast, during mid to late pollen development all three domains strongly enhanced promoter activity. These data suggest that *lat52* is transcriptionally regulated in a bi-phasic manner with domains A and B functioning during early development and domain C being utilised only at later stages of pollen development. However, in mature pollen domain A appears to be redundant in the presence of downstream sequences since domains B2, B3 and C together account for most of the promoter activity. Similarly, PBII alone activated -90 CaMV 35S specifically in pollen [36], but showed redundancy in the presence of downstream sequences. The independent

function of all three domains in mature pollen and the differential utilisation of domains A/B and C during development indicates that factors which bind A/B are present throughout, whereas factors binding within domain C only appear during mid to late stages of pollen development.

Organization of regulatory elements within domain B

Although PBII appears to be redundant in mature pollen, other pollen-specific regulatory elements are functional within sub-domains B2 and B3. These sub-domains are extremely AT-rich and contain no known pollen-specific regulatory sequences with the exception of a TCAC motif which is a reverse copy of the core pollen-specific 56/59 box [36]. However, this motif is unlikely to be functional since deletion of the TC nucleotides had little effect on *lat52* promoter activity [36]. Sub-domain B2 is 97% A+T and could therefore contain binding sites for high mobility group protein 1 (HMG 1)-like proteins. HMG1 enhances the sequence-specific DNA binding activity and transcriptional activation by the human HOXD9 protein [41]. HMG1-like proteins have been shown to bind to AT-rich regions in soybean nodulin promoters [14]. Such factors binding within sub-domain B2 could stimulate the binding of pollen-specific activators to the adjacent sub-domain B3.

Regulatory elements within domain C

Gain-of-function experiments defined domain C as an independent movable functional unit sufficient to direct pollen-specific transcription. These data confirm similar findings by Eyal *et al.* [7], but in our experiments enhancement was at least 100-fold lower. This apparent difference in activation potential most likely arises from the use of pJO64 (−64 CaMV 35S-luciferase) as the control plasmid by Eyal *et al.* which was reported to be inactive in pollen.

Within domain C the region −100 to −72, containing PBI was unable to activate CaMV 35S in pollen. In contrast, PBII (−194 to −176) was previously shown to activate CaMV 35S in a pollen-specific manner [36]. This indicates that despite their sequence identity PBI and PBII are likely to bind different transcription factors. It has been proposed that the *rbcS-3A* box II binding protein GT-1a could bind to PBI since mutation of PBI to that of a GT-1a binding site did not reduce promoter activity in pollen [7]. However, pairwise mutational analysis of PBI defined the central GG residues

as the only functional similarity between PBI and the GT-1a binding site (*rbcS-3A* boxII). Although PBI is unlikely to bind GT-1a, this does not rule out the possibility that it could be recognised by a closely related factor. Indeed, GT-1-related RT-PCR products were detected in tobacco pollen [7] and GT-1-like factors extracted from leaves have been shown to interact with non light-responsive promoters including the tobacco *PR-1a* [4] and the pollen-specific *NTP303* promoters [13].

Targeted mutational analysis within domain C led to the identification of the novel regulatory elements GAAA and TCCACCATA, which together with PBI form a strong pollen-specific transcriptional activator unit. Further gain of function and targeted mutational analyses demonstrated that PBI/GAAA and GAAA/TCCACCATA each could form pollen-specific but weaker activator units, but the combination PBI/TCCACCATA was inactive. These analyses also demonstrated that none of the three functional sequences could activate CaMV 35S alone. Two models for the action of putative transcription factors binding within domain C are presented in Figure 9. The first model involves a factor binding at the GAAA motif which facilitates and stabilises the binding of further transcription factors to GTGG (PBI) and TCCACCATA to contact TFIID and RNA polymerase II (Figure 9A). The second model involves factors binding to GTGG and TCCACCATA motifs which facilitate and stabilize the binding of a transcription factor to the GAAA motif (Figure 9B).

Identification of conserved pollen-specific regulatory elements within other late-pollen promoters would suggest a common regulatory pathway of gene expression during pollen development. Promoter analysis of the maize homologue of *lat52*, *Zm13*, showed that the region −314 to +61 was sufficient to direct pollen-specific expression [10] and that the proximal region −100 to −54 was essential for pollen specificity [11]. The GAAA motif which is essential for *lat52* proximal promoter activity occurs twice in reverse orientation within the *Zm13* proximal region at positions −93 and −74. The conserved position and extended sequence identity (AGAAATAGT) with sequences surrounding the GAAA motif within the *lat52* promoter (AGAAATAAT) suggests that the most proximal of the two *Zm13* GAAA motifs is likely to be functional. The GAAA motif is also present within the proximal promoter regions of the rice homologue of *lat52*, *PS1* (GGAAATCCT) [42] and the pollen-specific tobacco

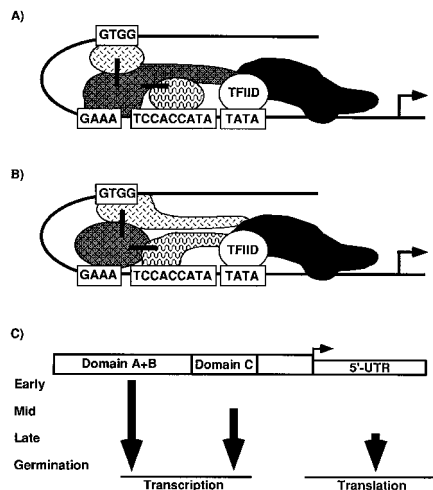


Figure 9. A, B. Functional regulatory elements and models for the interaction between putative transcription factors in mature pollen. C. Developmental role and mode of action of regulatory domains as defined here and by Bate *et al.* [2].

polygalacturonase gene *Npg1* (AGAAACTCT) [26] at positions –57 and –152 respectively.

In summary, data presented in this paper and in Bate *et al.* [2] have broadly defined the expression pattern of the tomato *lat52* gene into three phases during pollen maturation (Figure 9C). The first transcription phase (early to mid bicellular pollen) is regulated by the region –492 to –101 (domain A+B) and the second transcription phase (mid to late bicellular) is regulated by the region –101 to –52 (domain C). A third phase of developmental control (late bicellular to mature pollen) is regulated at the level of translation by the region +1 to +110 comprising the *lat52* 5'-UTR.

Acknowledgements

We are grateful to Colin Eady for providing the –492 *lat52-gus* transgenic pollen, Joy Wilkinson for providing the CaMV 53S-*gus* transgenic data and Graham Benskin for excellent greenhouse care. We gratefully acknowledge financial support for this work from BBSRC and the Royal Society as a Fellowship to D.T.

References

1. Batanero E, Villalba M, Lopez-Otin C, Rodriguez R: Isolation and characterization of an olive allergen-like protein from lilac pollen. Sequence analysis of three cDNA encoding protein isoforms. *Eur J Biochem* 221: 187–193 (1994).

2. Bate N, Spurr C, Foster G, Twell D: Maturation-specific translational enhancement mediated by the 5'-UTR of a late pollen transcript. *Plant J* 10: 613–623 (1996).
3. Bevan M: Binary *Agrobacterium* vectors for plant transformation. *Nucl Acids Res* 12: 8711–8721 (1984).
4. Buchel AS, Molenkamp R, Bol JF, Linthorst HJM: The PR-1a promoter contains a number of elements that bind GT-1 like nuclear factors with different affinity. *Plant Mol Biol* 30: 493–504 (1996).
5. Carrington JC, Freed DD: Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. *J Virol* 64: 1590–1597 (1990).
6. Eady C, Lindsey K, Twell D: Differential activation and conserved vegetative cell-specific activity of a late pollen promoter in species with bi- and tricellular pollen. *Plant J* 5: 543–550 (1994).
7. Eyal Y, Curie C, McCormick S: Pollen specificity elements reside in 30 bp of the proximal promoters of two pollen-expressed genes. *Plant Cell* 7: 373–384 (1995).
8. Gilmartin PM, Memelink J, Hiratsuka K, Kay SA, Chua N-H: Characterization of a gene encoding a DNA binding protein with specificity for a light-responsive element. *Plant Cell* 4: 839–849 (1992).
9. Green PJ, Yong MH, Cuozzo M, Kano-Murakami Y, Silverstein P, Chua N-H: Binding site requirements for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *rbcS-3A* gene. *EMBO J* 7: 4035–4044 (1988).
10. Guerrero FD, Crossland L, Smutzer GS, Hamilton DA, Mascarenhas JP: Promoter sequences from a maize pollen-specific gene direct transcription in tobacco. *Mol Gen Genet* 224: 161–168 (1990).
11. Hamilton DA, Roy M, Rueda J, Sindhu RK, Sanford J, Mascarenhas JP: Dissection of a pollen-specific promoter from maize by transient transformation assays. *Plant Mol Biol* 18: 211–218 (1992).
12. Hanson DD, Hamilton DA, Travis JL, Bashe DM, Mascarenhas JP: Characterization of a pollen-specific cDNA clone from *Zea mays* and its expression. *Plant Cell* 1: 173–179 (1989).
13. Hochstenbach R, de Groot P, Jacobs J, Schrauwen J, Wullens G: The promoter of a gene that is expressed only in pollen interacts with ubiquitous transcription factors. *Sex Plant Reprod* 9: 197–202 (1996).
14. Jacobsen K, Laursen NB, Jensen EO, Marcker A, Poulsen C, Marcker KA: HMG I-like proteins from leaf and nodule nuclei interact with different AT motifs in soybean nodulin promoters. *Plant Cell* 2: 85–94 (1990).
15. Jefferson RA, Kavanagh TA, Bevan MW: GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* 6: 3901–3907 (1987).
16. Kuhlemeier C, Cuozzo M, Green PJ, Goyvaerts E, Ward K, Chua N-H: Localisation and conditional redundancy of regulatory elements in *rbcS-3A*, a pea gene encoding the small subunit of ribulose-bisphosphate carboxylase. *Proc Natl Acad Sci USA* 85: 4662–4666 (1988).
17. Lombardero M, Barbas JA, Moscoso del Prado J, Carreira J: cDNA sequence analysis of the main olive allergen, *Ole e 1*. *Clin Exp Allergy* 24: 765–770 (1994).
18. McCormick S, Smith A, Gasser C, Sachs K, Hinchey M, Horsch R, Fraley R: The identification of genes specifically expressed in reproductive organs of tomato. In: Nevins D, Jones R (eds) *Tomato Biotechnology*, pp. 255–265. Alan R. Liss, New York (1987).

19. McCormick S, Twell D, Vancanneyt G, Yamaguchi J: Molecular analysis of gene regulation and function during male gametophyte development. In: Jenkins G, Schuch W (eds) *Molecular Biology of Plant Development*. Proceedings of SEB symposium No. 45. Company of Biologists, Cambridge (1991).
20. Muschietti J, Dircks L, Vancanneyt G, McCormick S: LAT52 protein is essential for tomato pollen development: pollen expressing antisense LAT52 RNA hydrates and germinates abnormally and cannot achieve fertilization. *Plant J* 6: 321–338 (1994).
21. Ow DW, Wood KV, DeLuca M, DeWet JR, Helinski DR, Howell SH: Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science* 234: 856–859 (1986).
22. Pe ME, Frova C, Colombo L, Binelli G, Mastroianni N, Tarchini R, Visioli G: Molecular cloning of genes expressed in pollen of *Sorghum bicolor*. *Maydica* 39: 107–113 (1994).
23. Perisic O, Lam E: A tobacco DNA binding protein that interacts with a light-responsive box II element. *Plant Cell* 4: 831–838 (1992).
24. Rodriguez-Garcia MI, Fernandez MC, Alche JD: Immunocytochemical localisation of allergenic protein (*Ole e I*) in the endoplasmic reticulum of the developing pollen grain of olive (*Olea europaea* L.). *Planta* 196: 558–563 (1995).
25. Schrauwen JAM, de Groot PFM, van Herpen MMA, van der Lee T, Reynen WH, Weterings KAP, Wullems GJ: Stage-related expression of mRNAs during pollen development in lily and tobacco. *Planta* 182: 298–304 (1990).
26. Tebbutt SJ, Lonsdale DM: Deletion analysis of a tobacco pollen-specific polygalacturonase promoter. *Sex Plant Reprod* 8: 242–246 (1995).
27. Topfer R, Matzeit V, Gronenborn B, Schell J, Steinbiss H-H: A set of plant expression vectors for transcriptional and translational fusions. *Nucl Acids Res* 15: 5890 (1987).
28. Tupy J, Ribhova L, Zarsky V: Production of fertile tobacco pollen from microspores in suspension culture and its storage for in situ pollination. *Sex Plant Reprod* 4: 284–287 (1991).
29. Turner R, Bate N, Twell D, Foster G: *In vivo* characterisation of a translational enhancer upstream from the coat protein open reading frame of potato virus S. *Arch Virol* 137: 123–132 (1994).
30. Twell D: The diversity and regulation of gene expression in the pathway of male gametophyte development. In: Scott RJ, Stead AD (eds) *Molecular and Cellular Aspects of Plant Reproduction*, pp. 83–115. Cambridge University Press, Cambridge (1994).
31. Twell D: Diptheria toxin-mediated cell ablation in developing pollen: vegetative cell ablation blocks generative cell migration. *Protoplasma* 187: 144–154 (1995).
32. Twell D, Patel S, Sorenson A, Roberts M, Scott R, Draper J, Foster G: Activation and developmental regulation of an *Arabidopsis* anther-specific promoter in microspores and pollen of *Nicotiana tabacum*. *Sex Plant Reprod* 6: 217–224 (1993).
33. Twell D, Wing R, Yamaguchi J, McCormick S: Isolation and expression of an anther-specific gene from tomato. *Mol Gen Genet* 217: 240–245 (1989).
34. Twell D, Klein TM, Fromm ME, McCormick S: Transient expression of chimeric genes delivered into pollen by microprojectile bombardment. *Plant Physiol* 91: 1270–1274 (1989).
35. Twell D, Yamaguchi J, McCormick S: Pollen-specific gene expression in transgenic plants: Coordinate regulation of two different tomato gene promoters during microsporogenesis. *Development* 109: 705–713 (1990).
36. Twell D, Yamaguchi J, Wing RA, Ushiba J, McCormick S: Promoter analysis of three genes that are coordinately expressed during pollen development reveals pollen-specific enhancer sequences and shared regulatory elements. *Genes Devel* 5: 496–507 (1991).
37. Ursin VM, Yamaguchi J, McCormick S: Gametophytic and sporophytic expression of anther-specific genes in developing tomato anthers. *Plant Cell* 1: 727–736 (1989).
38. Villalba M, Batanero E, Lopez-Otin C, Sanchez LM, Monsalve RI, Gonzalez de la Pena MA, Lahoz C, Rodriguez R: The amino acid sequence of Ole e I, the major allergen from olive tree (*Olea europaea*) pollen. *Eur J Biochem* 216: 863–869 (1993).
39. Willing RP, Bashe D, Mascharenhas JP: An analysis of the quantity and diversity of messenger RNAs from pollen and shoots of *Zea mays*. *Theor Appl Genet* 75: 751–753 (1988).
40. Willing RP, Mascharenhas JP: Analysis of the complexity and diversity of mRNAs from pollen and shoots of *Tradescantia*. *Plant Physiol* 75: 865–868 (1984).
41. Zappavigna V, Falciola L, Citterich MH, Mavilio F, Bianchi ME: HMG1 interacts with HOX proteins and enhances their binding and transcriptional activation. *EMBO J* 15: 4981–4991 (1996).
42. Zou J-T, Zhan X-Y, Wu H-M, Wang H, Cheung AY: Characterisation of a rice pollen-specific gene and its expression. *American Journal of Botany* 81: 552–561 (1994).