Participation of Chromatin in the Regulation of Phaseolin Gene Expression

TIMOTHY C. HALL, GUOFU LI, and MAHESH B. CHANDRASEKHARAN
Institute of Developmental and Molecular Biology and Department of Biology, Texas A&M University, College Station, TX77843-3155, USA

Received October 12, 1997 - Accepted November 27, 1997

Summary

It has long been recognized that multiple transcription factors contribute to the overall regulation of gene expression. Since it is inconceivable that a specific transcription factor is assigned to each protein-encoding gene, it has been suggested that several factors interact in various combinations (Goldberg, 1988; Thomas, 1993). However, it is not clear if such combinations are established by the relative linear positions of the corresponding cis-element, by the order in which binding sites are occupied (providing alternative surfaces for protein-protein interactions), or by both of these parameters, possibly in association with other influences. One of these influences is undoubtedly chromatin. While this has previously considered being only a repressive structure, the recent finding that several transcriptional regulators can acetylate histones is rapidly leading to the conclusion that modification of histones, and thus chromatin, is a vital step in activation of the transcriptional machinery (Wade and Wolffe, 1997).

Evidence for replacement of the repressive nucleosome structure by TFIID during transcription activation has been postulated by Van Holde and Zlatanova (1996) and may well be involved in phas promoter activation. Discerning actual steps by which the repressive nucleosome structure is modified or displaced will provide new insight to plant gene regulation, and it is tempting to think that histone acetylation will be part of the key. Once the chromatin structure is relaxed, transcription factors can gain access to the promoter, and a careful evaluation of changes in cis-element occupancy during embryogenesis using the sensitive DMS footprinting and LMPCR analysis technique will give important clues to the types of proteins involved. Another future opportunity for these approaches will be to explore events that lead to the cessation of expression as the embryo matures.

Key words: Phaseolin, in vivo footprinting, ligation-mediated PCR, nucleosome phasing.

Abbreviations: DMS = dimethylsulfate; LMPCR = ligation-mediated polymerase chain reaction; phas = phaseolin; TBP = TATA binding protein; TFIID = transcription factor IID.

Introduction

The investigation of seed protein expression is of great fundamental and practical importance. Agriculture is ultimately based on seed function and grain is typically the marketable crop product. Improvement of grain quality for nutrition and processing was an early goal of recombinant DNA technology, one that will be brought to practical reality over the next few years. From a basic science viewpoint, the dramatic contrast in spatial expression of many seed storage protein genes at very high levels during embryogenesis and their complete silence in vegetative tissues provides an excellent opportunity to explore the molecular basis of transcriptional regulation.

Intrinsic to gene expression is the establishment of the transcription complex. Early studies identified several sequence elements common to seed protein promoters. However, several of these elements, for example the G-box (CACGTG), are associated with the expression of many plant (and animal) genes. Indeed, promoters for seed protein genes are typically highly seed-specific, but since they can interact
Chromatin and Phaseolin Gene Expression

Fig. 1: Phas 5' promoter showing major regions and sequence motifs predicted on the basis of sequence homology.

- CCAAT box
- TATA box
- Various motifs:
  - CCACCA; NRS box - AG-1 box (AGAAAAA-GRCAA);
  - Anthocyanin (TAACTG);
  - CACA box (HAACACAWH>);
  - whHistoneBP - CaMV root (T-GACGT)
  - CCACM; CPRF site (ACGTGGy);
  - Octamer/Glutelin box (ATGCAGAA); Prolamin box (TGHAAARK).
  - Vicilin box (GCCACCTC); GAGA box (AGAGAGA); CCAC box (CCAC)

Core of matrix attachment region; em/en- embryo/endosperm

Multiple cis-elements are present in the phas promoter

The phas promoter has an array of motifs identical to consensus sequences of cis-elements known to interact with transcription factors (Fig. 1). The basal promoter, encompassing some 109 bp upstream of the transcription start site, contains a CCAAT element, three TATA boxes and several other cis-motifs. The region between -109 and -227 has 10 cis-motifs and transgenic tobacco plants containing 227 bp of the phas promoter fused to the uidA reporter (-227 phas/uidA) express low levels of GUS in embryo and endosperm tissues. The 68 bp region between -227 and -295 contains five sequence elements thought to be important for gene expression: three RY repeats, a G-box and a GC-rich motif. Although purine-pyrimidine-rich (legumin, RY or Sph) elements are often closely associated with seed-specific genes, no factor binding was reported prior to the recent findings of Suzuki et al. (1997) for a truncated version of VP1. The 68 bp region functions as a strong enhancer, and its placement in different orientations in transgenic tobacco alters the relative levels of GUS expression in embryo and endosperm (Van der Geest and Hall, 1996). Additional elements exist in the upstream region of the phas promoter. These include an AT-rich element that is important in yielding higher levels of expression from -795 constructs over -295 constructs (Bustos et al., 1989, 1991).

Materials and Methods

Nuclei isolation

Tissue was powdered in a mortar and pestle using liquid nitrogen, then treated with nuclei isolation buffer NIB 1: 0.5 M hexylene glycol, 20 mM KCl, 20 mM PIPES pH 6.5, 0.5 mM EDTA, 0.4 % Triton X-100, 0.05 mM spermine, 0.125 mM spermidine, 7 mM 2-mercaptoethanol, 0.5 mM phenylmethylsulfonyl fluoride and 0.5 % (v/v) Aprotinin (Sigma). The extract was filtered through 500, 250, 80, 45 and 20 μm mesh sieves and loaded onto a 20–80 % Percoll (Sigma) step gradient. For leaves, chloroplasts remained in the top layer whereas for seeds, starch grains pelleted through the 80 % Percoll layer. Nuclei were removed from the 20% 80 % Percoll interface, washed in NIB2 (NIB 1 without triton X-100), and stained with acridine orange for analysis under a microscope. We have found that using PIPES, pH 6.5, in place of HEPES, pH 7.4, dramatically improved the quality of the nuclei and routinely yielded 4×10⁷ nuclei from 5 g of leaf tissue and 5×10⁷ nuclei from 4 g of immature seeds (obtained from 40 seed pods).

LMPCR

LMPCR was done according to Mueller and Wold (1989) with the following modification. After ligation of the linker and ethanol precipitation, DNA was dissolved in 85 μL of H₂O. Subsequently, 10 μL 10 × PCR buffer (10 mM Tris HCl pH 8.0, 50 mM KCl,
**Results and Discussion**

The use of in vivo footprinting to assess cis-element functionality

Given the multiplicity of cis-elements within the context of the 5' phas promoter element, the question arises as to whether some or all are actually involved in transcriptional expression. While promoter truncation and permutation analyses are helpful in addressing this question, it is very difficult to assess combinatorial effects of these mutations; that is, if one cis-element is mutated to prevent factor binding, what effect(s) does this have on binding and assembly of other transcription factors? In contrast, in vivo footprinting using the small methylating agent dimethyl sulfate (DMS) can provide valuable insight to these questions (Hornstra and Yang, 1993; Lu et al., 1996). In this procedure, intact tissues (e.g. immature seeds or leaves) are exposed to DMS, which preferentially methylates G residues (and A residues to a much lesser extent) that are not protected by intimate interaction with a protein. Since no extraction or other perturbation is involved, the modification exerted by this agent reflects the in vivo status of the promoter at the time of treatment. Subsequently, DNA is extracted and subjected to piperidine treatment, which cleaves the methylated residues. We have recently used ligation-mediated PCR (LMPCR; Mueller and Wold, 1989) as a sensitive technique to determine the status of phas sequences in leaves and seed of transgenic tobacco containing a single copy of a chimeric – 1470phas/uidA construct (Li et al., 1998). This analysis has permitted evaluation of transcription factor occupancy in leaf and seed tissue of putative cis-elements identified primarily on the basis of their sequence similarity to elements known to be involved in transcriptional activity in other promoters.

Among the interesting findings from the in vivo footprinting data (G. Li and T. C. Hall, in prep.) is the fact that most of the sequences having a consensus to transcriptional motifs are in fact occupied in transcriptionally active seed tissues. This suggests that regulation of phas promoter expression requires the interaction of multiple factors in seed tissues. While many of these factors are undoubtedly involved in basic transcriptional events, it is likely that others participate in modulation of expression in ways that are yet to become evident. RY repeats have long been thought to be important for transcriptional activation in seed tissues (Bäumlein et al., 1992), and the in vivo footprinting approach shows clear evidence for protection of several RY elements in the phas promoter during embryogenesis. Mutagenesis of these elements is underway, and their effects on function and factor occupancy at other sites will be evaluated in transgenic plants. In contrast to the many discrete footprints evident in seed tissues, the phas promoter region in vegetative tissues appears to be devoid of factor interaction.

**Stringent spatial regulation of expression from the phas promoter**

The phas promoter provides an exceptional example of stringent spatial control. Phas promoter fusions to the uidA reporter gene fail to express β-glucuronidase (GUS) in any vegetative tissues, even in disorganized calli when treated with a variety of compounds known to be capable of stimulating transcriptional expression (Frisch et al., 1995). An even more stringent example is the ability to regenerate morphologically normal tobacco plants transgenic for phas-DT-A diphtheria toxin A-chain since a single molecule of DT-A is known to be lethal to a plant cell (Czako and An, 1991). Indeed, protoplasts were isolated from leaves of a tobacco plant transgenic for – 1470phas-DT-A in similar numbers to those obtained from wild type plants and regenerated to yield phenotypically normal plants (Li and Hall, unpubl.). While the above examples demonstrate complete lack of activity of the phas promoter in vegetative tissues, it is highly active after the mid-maturation stage during embryogenesis. These differences in spatial expression could be due to the absence of transcription factors needed for phas activation in vegetative tissues. However, when identical phas/uidA constructs were supplied as naked DNA for bombardment of tobacco or bean leaves (or electroporation of protoplasts), expression was readily obtained. This represents a striking difference in expression from naked and chromosomally integrated DNA. These experiments reveal that factors capable of trans-activating expression from the phas promoter are present in leaf tissue, resulting in transient expression. Together with the complete lack of expression from stably integrated phas constructs this supports the notion that, as described below for animal systems, chromatin conformation is an essential component of spatial regulation of phas transcription as well.

**The involvement of chromatin structure in phas gene regulation**

Nucleosome spacing is not involved in phas spatial regulation

Chromatin controlled regulation of expression is well documented for several eukaryotic genes from yeast and animals (Simpson, 1991; Wolffe and Dimitrov, 1993). Travers (1992) has suggested that transcriptional competence may be accompanied by chromatin rearrangement, and it was tempting to consider the possibility that the differences in nucleosome spacing observed by Murray and Kennard (1984) for leaf and developing cotyledons of Phaseolus vulgaris reflected some type of reprogramming associated with transcriptional acces-
Fig. 2: Nucleosomal repeats in tobacco leaf and seed nuclei. Aliquots equivalent to 2 million nuclei were treated with MNase (units used are indicated above the lanes). After DNA extraction, ~2 μg of DNA from each sample was analyzed on a 1.5% agarose gel. M, marker DNA (1 kb ladder, BRL).

Nucleosome positioning refers to the identical placement of histone octamers along a particular DNA sequence in all cells of a given population (Simpson, 1991). Precise positioning of nucleosomes has been shown to be a major determinant in the expression of several genes, including mouse mammary tumor virus (MMTV) (Archer et al., 1991) and yeast STE6 (Shimizu et al., 1991). To investigate whether nucleosome positioning plays a role in phas gene regulation, we used DNase I footprinting in vivo to probe the chromatin structure of both active and inactive state of the phas promoter (Li et al., 1998). Our results show that a nucleosome is rotationally positioned on the TATA region of the inactive phas promoter. LMPCR analysis of the promoter following DNase I digestion of leaf nuclei revealed a clear 10 bp pattern of protection and cleavage. This is characteristic for DNA rotationally positioned on a nucleosome as bases adjacent to the histone core are protected from cleavage and those facing away from the core are preferentially cleaved. Further support for this interpretation came from a comparison of cleavage for the bottom strands of the phas promoter (Fig. 3), since cleavage was staggered by 2–4 bp compared with top strand (Li et al., 1998). The 10 bp repeat pattern on the phas promoter in leaf tissue extended over some 150 bp, consistent with DNA being wound twice round the histone core, as is typical for a nucleosome. Interestingly, the three TATA elements...
are protected in DMS footprinting of seed tissue, showing that they are in phase, each facing towards the histone core when transcriptionally inactive in vegetative tissue (Fig. 3).

Potential role of DNA topology in phas gene regulation

The ability of the phas promoter sequence to position nucleosome(s) is exciting as there are only a few systems for which such ability has been described (Lu et al., 1994), none of them from a plant promoter. It will be valuable to determine what feature or features of the phas promoter are responsible for positioning. One aspect is likely to be DNA topology. While DNA is typically represented as a linear helix, differences in tilt and roll angles between the individual base-pairs give rise to curvature that is intrinsically dependent on sequence (Sinden, 1994). Application of the program CURVATURE (Shpigelman et al., 1993) to the phas promoter reveals that it has considerable intrinsic curvature. One region is adjacent to the phased TATA elements, and it is tempting to think that this region may participate in the initial setting of the nucleosome position.

Another potential role for DNA topology in gene regulation is the context it provides for cis-elements in transcriptionally active tissues. As was mentioned earlier, transcription factors (G-box binding factors) binding to G-box (CACGTG) motifs have been implicated in the expression of several tissue-specific genes that have different spatial expression (Menkens et al., 1995). One way by which discrimination for a specific member, or at least a subset of members, of a given family of transcription factors may be achieved could be the topological environment of the sequence motif, for example, the G-box motif may lie within a linear region of DNA, or in a region of positive or negative curvature. Future experiments in which mutations are designed to alter the topological environment of important cis-elements will permit evaluation of the contribution of intrinsic curvature to phas gene expression.

Nucleosome displacement is associated with active transcription in seed tissues

Upon transcription activation, the nucleosome over the TATA region of the phas promoter is disrupted. Although DNase I footprinting of the phas promoter in nuclei from seed tissues produced a 10 bp repeat cleavage profile (Fig. 3), it only extended for some 50 bp. This, in combination with data from DMS footprinting in vivo, suggested that the short extent of 10 bp repeat reflects the presence of the histone octamer-like substructure within TFIIID rather that the normal histone octamer (Li et al., 1998). Indeed, the pattern is similar to that found by Xie et al. (1996) for the Ad2ML promoter when bound by TFIIID, which was thought to represent protection by the histone-like domain of that factor. This possibility has been supported experimentally by the demonstration in vivo that TFIIA interacts with the phas promoter DNA to give a similar short stretch of 10 bp repeat cleavages (Li et al., 1998). In vivo MNase footprinting provides additional indirect evidence supporting this alternative. High resolution analysis of MNase digestion products revealed different patterns for leaf and seed nuclei. In leaf nuclei, the prominent cleavage site at +9, and many other sites extending over a region of more than 145 bp, are protected compared with naked DNA, consistent with the presence of a nucleosome (Fig. 4). In seed nuclei, although the prominent cleavage sites adjacent to +9 are protected, no protection is evident upstream of -40. Furthermore, sites near -40, -18 and +40 were hypersensitive to MNase digestion of chromatin. The generally protected region from -18 to +40 corresponds with the concept that the DNA is wrapped around the histone octamer-like substructure within TFIIID, but is much shorter than that expected to be protected by a nucleosome.

Higher order chromatin structure may also modulate expression from phas

The data discussed above reveal that the nucleosome positioned on the TATA region of phas promoter plays a decisive role in the lack of expression of phas gene in vegetative tissues. However, the contribution of chromatin on phas gene expression is not always negative. Regions 5' and 3' flanking the phas gene are extremely A/T-rich, contain several motifs commonly found in MARs and binds strongly to isolated to-
bacco matrices (Van der Geest et al., 1994). The 5' MAR is centered near position -800 and the 3' MAR at approximately +2500, suggesting that the phas gene lies on a chromatin loop of only 3.3 kb (Van der Geest et al., 1994). This arrangement contribute to the extremely high and position-independent levels of expression from the phas promoter in developing seeds of transgenic plants and demonstrates a positive role for chromatin structure in gene expression.

A proposed model for regulation of phas expression

It is known that the 5' promoter region of phas contains many motifs corresponding to both tissue-specific and tissue nonspecific transcription factor binding sites (Fig. 1; Bustos et al., 1991). When present as naked DNA, these motifs are capable of interacting with any suitable transcription factor, permitting recruitment and assembly of the basal transcription complex in the cell into which they are introduced (Fig. 5 a). However, when the phas promoter is present as chromatin, a nucleosome is rotationally positioned over the TATA box region, completely silencing the phas promoter (Fig. 5 b). During embryogenesis, histone octamers including the one over the TATA region must be remodeled or removed before interaction between cis elements and trans-acting factors can occur. Histones on the phas promoter can be remodeled in several different ways. Remodeling may involve plant homologs of known nucleosome remodeling factors such as SWI/SNF complex (Côté et al., 1994). Alternatively, remodeling may operate through histone acetylase activity (Wade and Wolffe, 1997). In addition, nucleosomes on the phas promoter can be removed through pre-emptive competition which requires DNA replication for disruption of chromatin structure, after which a trans-acting factor can bind to prevent reassembly of the newly replicated DNA into chromatin (Felsenfeld, 1992). Endoduplication of cotyledonary DNA, necessary for the pre-emptive competition model, has been observed in Phaseolus (Bino et al., 1993). It is also possible that the phas promoter is activated by combinations of the above mechanisms. During late embryogenesis, the phas promoter is again wrapped around histone octamer, returning to silenced state.

Acknowledgements

We thank Alan Wolffe for discussions and National Science Foundation grant MCB-9604617 for support of some of the work reported here.
References


