Plant Science 217–218 (2014) 109–119

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Review

Identification and validation of promoters and *cis*-acting regulatory elements

Carlos M. Hernandez-Garcia, John J. Finer*

Department of Horticulture and Crop Science, OARDC/The Ohio State University, 1680 Madison Avenue, Wooster, OH 44691, USA

ARTICLE INFO

Article history:
Received 4 September 2013
Received in revised form 3 December 2013
Accepted 5 December 2013
Available online 14 December 2013

Keywords:
Transcription
Genetic engineering
Synthetic promoters
Chromatin structure
Transcriptomics
Gene expression systems

ABSTRACT

Studies of promoters that largely regulate gene expression at the transcriptional level are crucial for improving our basic understanding of gene regulation and will expand the toolbox of available promoters for use in plant biotechnology. In this review, we present a comprehensive analysis of promoters and their underlying mechanisms in transcriptional regulation, including epigenetic marks and chromatin-based regulation. Large-scale prediction of promoter sequences and their contributing *cis*-acting elements has become routine due to recent advances in transcriptomic technologies and genome sequencing of several plants. However, predicted regulatory sequences may or may not be functional and demonstration of the contribution of the element to promoter activity is essential for confirmation of regulatory sequences. Synthetic promoters and introns provide useful approaches for functional validation of promoter sequences. The development and improvement of gene expression tools for rapid, efficient, predictable, and high-throughput analysis of promoter components will be critical for confirmation of the functional regulatory element sequences identified through transcriptomic and genomic analyses.

Contents

* Corresponding author. Tel.: +1 330 263 3880; fax: +1 330 263 3887.
E-mail address: finer.1@osu.edu (J.J. Finer).

0168-9452
http://dx.doi.org/10.1016/j.plantsci.2013.12.007

1. Introduction

Multicellular organisms maintain the same inherited genetic material in most of their cells throughout different cellular and developmental processes. However, organisms have acquired a diverse array of molecular mechanisms including coordinated expression of genes that tightly control metabolism during differentiation and development. Regulation of gene expression in different tissues and organs, during various growth and developmental stages, or as a consequence of external stimuli is mediated at the transcriptional, post-transcriptional, and post-translational level. Transcriptional regulation plays the greatest role in the activation and suppression of expression, and is largely controlled through gene promoters and their contributing *cis*-acting elements [1].

In the simplest terms, gene promoters are DNA sequences located upstream of gene coding regions and contain multiple *cis*-acting elements, which are specific binding sites for proteins involved in the initiation and regulation of transcription. DNA sequences located in the 3′-flanking or downstream region of the transcribed region, or even within the transcribed region, can also influence the initiation of transcription; however, these gene regulatory elements will not be discussed here. Promoters of protein-encoding genes often contain a "core promoter", which is a region located ~40 bp upstream of the transcriptional initiation site and comprises the TATA box [2]. The TATA box is the binding site for the transcription initiation factor TFIID TBP (TATA-box-Binding Protein) subunit. The core promoter also contains *cis*-elements that are binding sites for the basic transcriptional machinery, including RNA polymerase II and its corresponding subunits (Fig. 1) [3]. A protein complex, including general transcription factors such as TFIID and TFIIB is formed with RNA polymerase II prior to initiation of transcription [4].

Upstream of the core promoter region are the proximal and distal regions of promoters. Proximal and distal regions of the promoter contain different regulatory sequences such as enhancers, silencers, insulators, and *cis*-elements that contribute to the fine regulation of gene expression at the transcriptional level (Fig. 1) [4]. The physical demarcation of the upstream regions that contribute to the "full promoter" is more fluid than that of the core promoter, and the size of the active, fully functional promoter depends directly on the positional and combinatorial understanding of the *cis*-acting elements present in both the proximal and distal regions. During transcription, co-activators and transcription factors bind to specific DNA motifs and simultaneously interact with the transcriptional machinery attached to the core promoter [4]. This complex DNA/protein interaction leads to the activation, enhancement, or suppression of transcription. Thus, regulation of transcription depends on the: (a) availability and activity of transcription factors, and (b) the type, number, position, and combination of regulatory elements present in and around the promoter [1].

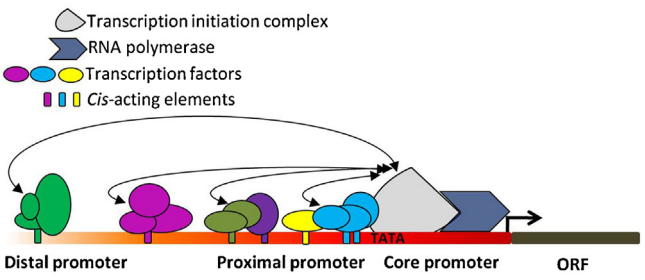

Fig. 1. Simplified model of transcriptional regulation of protein-encoding genes.

Regulation of gene expression at the promoter level is mainly controlled by the *cis*-acting elements localized upstream of the transcriptional start site. The physical interaction between regulatory proteins and the basic transcriptional machinery is straightforward during initiation of transcription due to the location of proximal elements to the core promoter. Distal promoter elements, located far away from the transcriptional start site can also impact gene expression. The mechanisms of how distal elements come into close proximity to the core promoter to modulate gene expression during transcription involve DNA folding mediated by conformational changes in the 3-dimensional structure of DNA and chromatin.

2. Chromatin structure and its role during transcription

Linear models depicting gene promoters (Fig. 1) are often simple representations of the contributing *cis*-acting elements positioned upstream of the transcriptional start site of a gene. However, *in vivo* regulation of transcription is more dynamic and portrays an augmented level of complexity. The 3-dimensional organization of DNA, influenced by folding and the association with chromatin appears to be highly organized and allows *cis*-acting elements located in distant regions to fold and spatially become proximal to the regulatory complex (Fig. 2) [5]. Under this dynamic definition of promoters, introns, 3′UTR, 5′UTR and even regulatory sequences positioned up to several Kb, and in extreme cases, more than 1 Mb away from the core promoter, can influence transcription rates [6].

Compacted assembly of the genomic DNA, wrapped by histones in a small nucleus is also a major constraint for transcriptional regulatory proteins and RNA polymerases limiting access to DNA and, thus, leverages gene transcription. Chromatin-based gene regulation includes replacement of common histones with specialized variant types and total or partial removal of histones from DNA [7]. The hypermobile animal nucleosomes containing the H3.3 and H2A.Z histone variants are relatively unstable, with these histones being easier to displace from DNA. These histones are predominantly associated with promoter regions, enhancers, and gene coding regions, where the nucleosomes are disrupted and reformed rapidly during transcription [8].

Post-translational modification of histones plays an important role in regulating transcription. Histones can be modified at their N-terminal tails through the addition of functional groups including methyl, acetyl, and phosphoryl. The addition of epigenetic marks to histones leads to activation or silencing of transcription as a result of either loosening or enhancing the association between histones and DNA [9]. For example, the trimethylation of the H3 histone protein at lysine 4 (H3K4me3) or lysine 27 (H3K27me3) are well-studied in animals and plants. The H3K4me3 histone variant is highly represented near the 5′ end of actively transcribed genes and associated with transcriptional initiation of the *Flowering Locus C* (FLC) in *Arabidopsis* [10]. Contrarily, the H3K27me3 variant is linked to gene silencing *via* chromatin condensation during plant development processes [9]. Transcription factors also have the ability to recruit coactivator proteins that acetylate histones and, thus, positively affect the activation of transcription.

Studies of chromatin-based regulation in plants are emerging and yet more research on epigenetic features and genome-wide mapping of histone modifications is required to predict active promoters and enhancers, and gain a better understanding of regulation of transcription at the promoter level. Transcription of transgenes in transformed organisms is also subject to chromatin-based regulation, albeit the degree of regulation depends on where the transgene is integrated. Most conventional DNA introduction methods result in somewhat random integration of transgenes in the host genome, including transcriptionally active and inactive

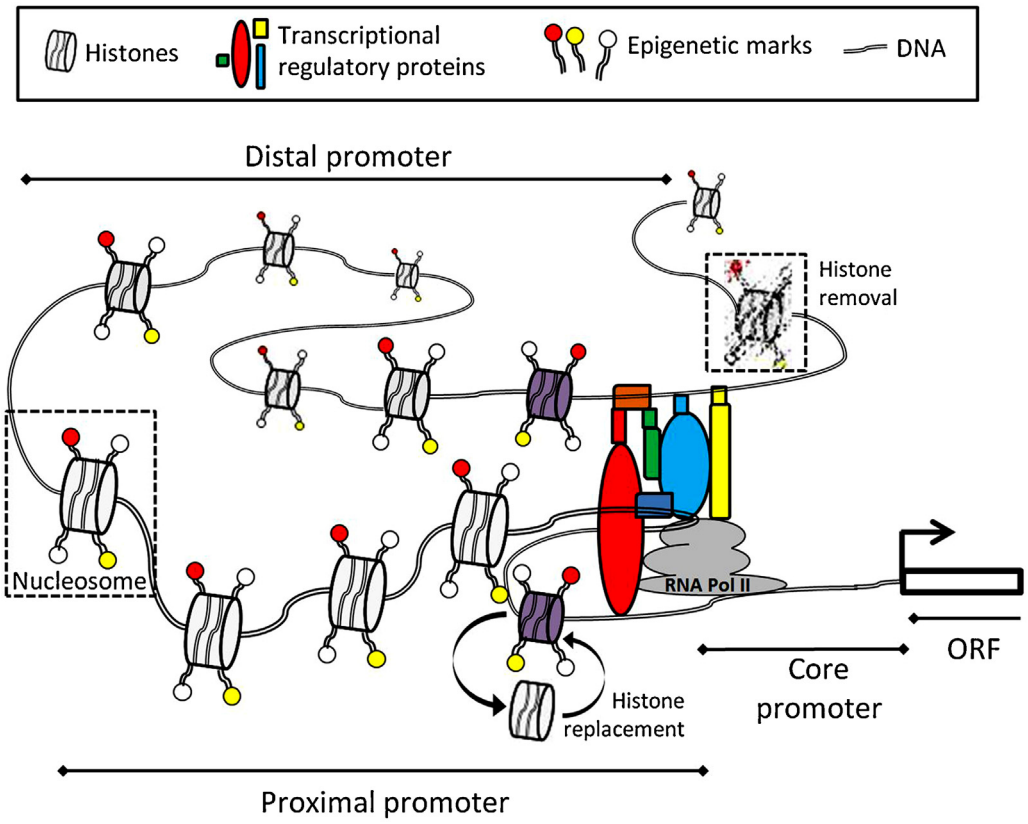

Fig. 2. Model of promoter showing core promoter, proximal promoter, and distal promoter regions. This model shows dynamic regulation of gene transcription, integrating the 3D structure of chromatin, epigenetic marks, and folding of the distal promoter sequences to become proximal to the transcriptional complex.

regions, which are often determined by chromatin status and epigenetic signature.

3. Plant promoters used in genetic engineering

Studies of gene promoters are central to understanding the global regulation of gene expression in plants. Isolated promoter sequences and their contributing elements are also critical for fine regulation of introduced transgenes, including protein-encoding genes and noncoding DNA sequences intended for gene silencing *via* RNAi-mediated silencing. Although isolated promoters can retain their native functionality in transgenic plants, qualitative and quantitative variations in transgene expression are not uncommon and should be carefully studied for each candidate promoter. The introduced transgenic DNA is affected by its location in the genome (position effect) and the genetic background of the transgenic organism.

Promoters used in biotechnology applications can be grouped into the following classes: (a) constitutive promoters – active in most of the tissues and developmental stages; (b) spatiotemporal promoters – provide tissue-specific or stage-specific expression; (c) inducible promoters – regulated by the application of an external chemical or physical signal; and (d) synthetic promoters – contain defined regulatory elements located adjacent to or within promoter sequences. Synthetic promoters can drive constitutive, spatiotemporal, inducible, and even unique combinations of transgene expression patterns, depending on the included elements.

3.1. Constitutive promoters

Constitutive promoters drive somewhat constant levels of gene expression in all tissues, at all times. However, few, if any promoters are truly constitutive. Most of the promoters that are classified as constitutive will display moderate expression in many tissues and higher expression in either rapidly growing meristematic tissue or vascular tissues. In reality, constitutive promoters are promoters that drive expression in most tissues under many different conditions. Constitutive promoters are often used when evaluating transgenes as transgene effects may be easier to score if the introduced gene can be expressed in most tissues under many different conditions. Constitutive promoters typically originate from either plant viruses or plant housekeeping genes.

Among the viral promoters, the CaMV35S promoter [11] from the cauliflower mosaic virus is one of the most widely utilized promoters for basic research and the development of transgenic plants. The constitutive expression of this promoter tends to be relatively high in different tissues of many plants. This strong expression in different plant tissues seems to result from an additive effect of multiple tissue-specific elements [12]. Site-specific mutation of some of the elements within the CaMV35S promoter resulted in a loss of expression in certain tissues or under certain conditions [12]. Viral promoters may have evolved by mutation of existing elements within their promoters or acquisition of elements from their host, evolving to become more successful in utilizing or hijacking the transcriptional machinery of their host. The simple model proposed for the CaMV35S promoter, where promoters are composed of strings of separate elements, each contributing to the total activity of the promoter, remains central to our conceptual understanding of promoter functionality. Unfortunately, many of the models for functionality of plant promoters are based on this CaMV35S model of assembled tissue-specific elements, which may not be totally valid for promoters from higher plants. Native plant constitutive promoters may not be a collection of numerous tissue-specific elements as with viral promoters [11,13] but may be composed of

non-specific elements that are simply more efficient at protein recruitment for transcription.

High-expressing housekeeping genes that encode abundant proteins required for basic functions in plant cells are a good source of strong native constitutive promoters. Strong constitutive promoters have been identified and isolated from the highly expressing *Ubiquitin*, *Actin*, *Tubulin*, and *EIF* (eukaryotic initiation factor) genes. Promoters from these genes are highly active in almost all organs and tissues and throughout most of the life cycle of a plant. Recently, the search for strong constitutive promoters has also expanded to other gene families encompassing the *APX* (ascorbate peroxidase), *PGD1* (phosphogluconate dehydrogenase), and *R1G1B* (R1G1 domain containing protein B) [14].

Unlike spatiotemporal and inducible promoters, few elements from endogenous plant constitutive promoters have been identified. A 38-bp poly(dA-dT) element in the upstream region of the rice *ACT1* promoter functions as a positive regular of gene expression; while, its CCCAA repeat contributes to complex tissue-specific expression [15]. A DNA fragment located −151 to −73 in the *NeIF4A-10* promoter directs high levels of expression although mutation of its predicted elements does not completely eliminate expression, suggesting a combinatorial effect of multiple elements or long element sequences [16]. We have identified a well conserved G-box like motif within strong constitutive polyubiquitin promoters, including the soybean Gmubi3 and Gmubi7, rice RUBQ2 and rubi3, maize Zmubi1, sunflower UbB1, switchgrass PvUbi1, and potato Ubi7 (Supplemental Table 1). Tetramers of this element gave exceptionally high expression levels of transient expression and in stably transformed soybean hairy roots (Fig. 3; Hernandez-Garcia, unpublished data), suggesting that the G-box motif contributes considerably to the strong expression mediated by polyubiquitin promoters.

3.2. Spatiotemporal promoters

Strong constitutive promoters are useful when high levels of transgene expression are continuously needed during the life cycle of a plant. Unlike constitutive expression of RNAi-encoding sequences, high and continuous overexpression of certain protein-encoding genes may be unnecessary and even harmful to plant cells. For instance, constitutive expression of genes associated with stress tolerance and defense responses may lead to unintended phenotypes, including altered plant growth and development, abnormal morphology, unexpected activation of defense pathways in the absence of pathogens, and fitness costs [17]. Examples of transcriptional regulators that show enhanced traits but accompanied by detrimental effects when over-expressed in a constitutively manner include the *DEHYDRATION ELEMENT BINDING* (*DREB1a*), *SALT TOLERANCE ZINC FINGER* (*STZ/ZAT10*), and *WAX PRODUCTION 1* (*WXP1*) [18].

Spatiotemporal promoters provide more precise control of native genes and transgenes, and restrict gene expression to certain cells, tissues, organs, or developmental stages. An ideal tissue-specific promoter will only be active in certain tissues leaving all other tissues unaffected. In the real world of transgenics, transgene expression regulated by a spatiotemporal promoter can be leaky, giving weak expression in tissues or at times that may be unexpected, based on expression of the native gene associated with the promoter. Differential expression driven by a tissue-enhanced promoter in transgenics and its native context would be in part attributed to the potential loss of *cis*-regulatory elements during promoter cloning, changes in chromatin-mediated regulation in transgenics, and the influence of native enhancers located in close proximity to the transgenic locus. Fine dissection of *cis*-acting elements present in spatiotemporal promoters is central to build promoters with more targeted and precise regulation of transgenes.

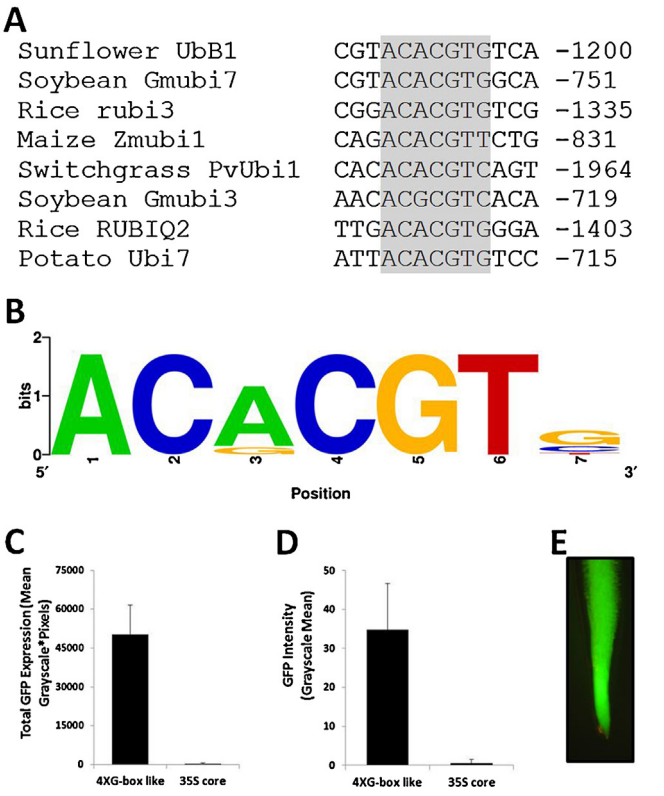

Fig. 3. Validation of a well-conserved G-box like motif present in strong polyubiquitin promoters (Hernandez-Garcia, unpublished data). (A) Sequence alignment. (B) Motif logo generated using WebLogo (http://weblogo.berkeley.edu/). (C)–(D) GFP expression mediated by tetramers of the G-box like motif cloned in front of the 35S minimal promoter (35S core) using transient expression (C; $n = 4–6$) and soybean hairy roots (D; $n = 22–23$). (E) Soybean hairy root showing high levels of GFP expression mediated by the 4xG-box like motif. The validation tools used for gene expression analysis were previously described [41,79].

A range of spatiotemporal promoters is available [19,20], although here we only focus on a few groups from this promoter class. Seed-specific promoters isolated from genes with restricted or enhanced expression during seed development are the most often reported spatiotemporal promoters. Seed-specific promoters have a wide range of applications including tissue-specific targeting of industrial and pharmaceutical compounds, and development of transgenic seeds with improved nutritional quality and better functional quality of milled grain. Promoters from cereal seed storage protein genes such as *Hordein* and *Glutenin* genes have been exploited for production of recombinant proteins in grains [21]. Similarly, several seed-specific promoters have been cloned from different plants (Supplemental Table 1).

Fruit specific promoters deserve special attention as they can lead to generation of fruits with improved post-harvest quality and enhanced nutritional value and agronomic quality through genetic engineering. Promoters that enrich gene expression in fruits can also be utilized to direct the production of antibodies, biopharmaceuticals and edible vaccines. *Expansin* genes encode plant cell-wall loosening proteins involved in cell-wall modification during fruit ripening [22]. Expansin promoters isolated from different plants showed specific expression in ripened fruits (Supplemental Table 1). Other promoters associated with fruit ripening such as the ACC-oxidase, E8 and PG (polygalacturonase) have also been used to direct expression in fruits (Supplemental Table 1).

Tubers and root storage organs from potato, sweet potato, yam, taro, and cassava are important staple food sources in many

developing countries; thus, tuber/storage organ-specific promoters are important for genetic engineering of these crops. The search for these promoters has focused on genes involved in the deposition of starch and storage of glycoproteins. Several promoters, including the dioscorin *pDJ3S*, potato class I patatin, GBSS-granule-bound starch synthase, sporamin, and ß-amylase have been reported as tuber/storage organ-specific promoters in cassava, carrot, potato, and sweet potato (Supplemental Table 1). These promoters often contain different sugar-responsive elements like the TGGACGG motif present in the sporamin and ß-amylase promoters [23]. SURE (Sucrose Responsive) elements were also found in sucrose-responsive patatin promoters [24]. Native regulation of genes is now feasible using engineered transcriptional regulators specifically designed to bind *cis*-regulatory elements within promoters [18], and native regulation of tuber/storage organ-specific genes would be a good approach to generate crops with enhanced agronomic traits and better nutrient content.

Anther- and pollen-specific promoters are useful to control male sterility, which is an important trait in plant breeding. Numerous anther- and pollen-specific promoters have been isolated from different plants, including the *RA8* promoter from rice, *A9* promoter from *Arabidopsis*, and *TA29* promoter from tobacco [19]. These promoters have been frequently used to regulate the expression of the barnase protein, which is a potently cytotoxic ribonuclease isolated from *Bacillus amyloliquefaciens* and optimized for expression in plants [25]. The use of pollen-specific promoters to control a cytotoxic protein in transgenic plants would mitigate the potential flow of transgenes into wild populations that reproduce by cross-pollination. Yet, the rate of pollen survival in these transgenics should be careful estimated and minimized; otherwise the introgression of a male sterility trait into wild populations can be ecologically disastrous.

3.3. Inducible promoters

Inducible promoters are responsive to environmental stimuli and provide precise regulation of transgene expression through external control. This group of promoters has a broad spectrum of potential applications to control expression in both small experimental settings and at a large agricultural scale by the application of chemical sprays. Inducible promoters are also useful for the regulation of potentially lethal genes or stress-related genes that are activated as a result of biotic and abiotic stresses. Inducible plant promoters can be usually classified as responsive to endogenous signals (plant hormones), external physical stimuli (biotic and abiotic stresses), and external chemical stimuli.

Numerous promoters responsive to hormones from plants, insects, and mammals have been characterized and effectively used in plants. For instance, chimeric systems containing the insect ecdysone receptor ligand-binding domain are inducible in plants using commercial insecticides containing either tebufenozide or methoxyfenozide [26,27]. Other promoters such as the glucocorticoid-mediated GVG and the estrogen-inducible XVE have been constructed using mammalian components and utilized for regulation of gene expression in plants [28,29]. The ethanol inducible *alc* system has also been widely used in plants. This system consists of two components: the ALCR transcription factor encoded by the *alcR* gene and the *alcA* promoter activated by ALCR in response to exogenous application of ethanol or acetaldehyde [30]. While expression of any transgene of interest is controlled by the *alcA*-inducible promoter, the expression of ALCR factor can be driven by any appropriate constitutive, inducible, or spatiotemporal promoter [31], facilitating the investigation of temporal and spatial gene activity. Inducible promoter systems using heterologous components from other organisms are advantageous as their activators typically do not induce endogenous pathways in plants. The most common applications for these systems include the increased production of recombinant proteins, functional analysis of lethal and harmful genes, and the production of marker-free transgenic plants utilizing targeted excision of selection markers.

During evolution, plants have acquired intricate mechanisms of gene regulation to mitigate the effect of adverse environmental conditions. Those genes, which are differentially expressed during stress, are good sources of stress-responsive promoters and *cis*-elements. The promoters of the rice *OsNCED3* and *Wsi18* genes, implicated in the synthesis and signaling of ABA, were highly-inducible after drought, ABA, and high-salinity treatments in transgenic rice [32,33]. The *Arabidopsis Rd29A* promoter was successfully used to mediate drought-specific expression of *DREB1A* in transgenic wheat [34].

The *cis*-regulatory elements present in stress-responsive gene promoters have also received special consideration as they provide insights into gene regulation and plant signaling under stress conditions. Examples of common stress-responsive elements comprise the dehydration-responsive element DRE (A/GCCGAC) implicated in the regulation of cold and dehydration responses in *Arabidopsis* [35], the low temperature responsive element C-repeat binding factor (CBF) [36], and the ABA responsive element ABRE (ACGTGG/T) that regulates dehydration and salinity responses in *Arabidopsis* and rice [37].

Defense-response promoters are also important in biotechnological applications. Although many genes conferring disease resistance have been discovered, the current toolbox of defense-response promoters may limit the development of resistant transgenic crops. The ideal pathogen-inducible promoter would be rapidly activated by a wide array of pathogens, and be inactive under disease-free conditions. Otherwise, the biosynthesis of abundant, unnecessary recombinant proteins controlled by strong constitutive promoters in transgenic plants can represent a high metabolic cost and eventually impact the energy allocated into traits of interest such as yield and biomass. Pathogen inducible promoters have been isolated from numerous genes associated with defense responses in plants, including defensin promoters [38], the rice OsPR10a promoter – induced by pathogens and defense response hormones [39], and the barley Germin-Like GER4 promoter – highly inducible after infection by biotrophic and necrotrophic pathogens [40]. This latter promoter contains multiple W-boxes with a typical TGAC core motif, which is a binding site for WRKY transcription factors and a motif required for pathogen-induced activity [40].

Mechanical wounding often mimics the physical aspect of pest activity and wound-responsive promoters can also be suitable for driving certain defense genes. Identification of wound-responsive promoters has typically been restricted to Pathogenesis-related (*PR*) genes [38]. However, other gene families such as bZIP, MYB, MYC, WRKY, and ERF transcription factors could also be good candidates for identifying wound-inducible promoters. Wound inducible promoters that mediate rapid gene responses would be best for controlling genes that provide protection during the initial stages of pathogen colonization, or immediately after arrival of insects. Late wound inducible promoters would also be good candidates to control genes that provide protection in late stages of pathogen invasion, which may facilitate a more comprehensive regulation of defense response genes. We previously identified 10 GmERF (*Glycine max Ethylene Response Factor*) promoters in soybean [41]. The GmERF3 promoter showed a delayed activation in response to wounding and may be either a late wound response promoter or senescence-associated promoter. This promoter contained a relatively large regulatory element (GGATTCAAGTTTAACC); promoter activation was minimized only when most of this element was modified (Hernandez-Garcia, unpublished).

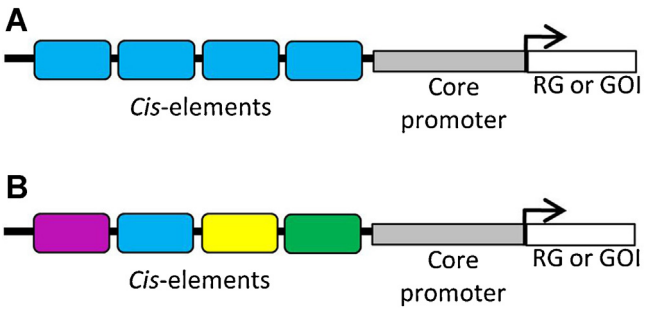

Fig. 4. Representative structure of synthetic promoters. (A) Tetrameric repeats of the same regulatory elements driving either a reporter gene (RG) or other gene of interest (GOI). (B) Example of a synthetic promoter-containing different regulatory elements controlling a reporter gene (RG) or a gene of interest (GOI).

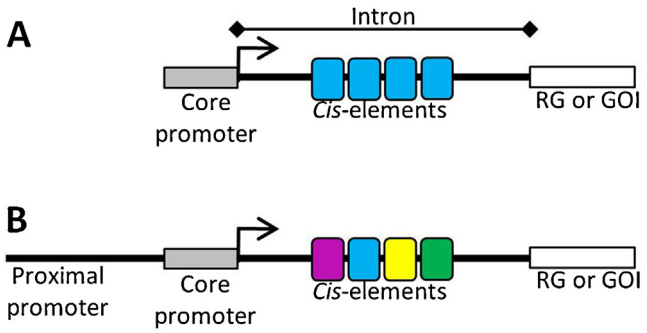

Fig. 5. A proposed methodology to improve regulation of gene expression using *cis*-acting elements located within intronic sequences. (A) Structure of an independent intron-containing tetrameric repeats of the same *cis*-element regulating a reporter gene (RG) or other gene of interest (GOI). (B) Structure of an intron-containing promoter harboring multiple copies of different *cis*-elements to mediate a reporter gene (RG) or other gene of interest (GOI).

3.4. Synthetic promoters

Synthetic promoters are composed of unique combinations of core promoter sequences, leading introns, and proximal and distal promoter sequences, which may all contain specific regulatory elements. Synthetic promoters differ tremendously from native promoters because they can provide expression profiles that do not exist in nature, ultimately improving the toolbox of useful promoters. Synthetic promoters are classically constructed by combinatorial engineering of *cis*-elements which include enhancers, activators or repressors directly upstream of the core promoter sequence (Fig. 4) [42]. This approach has been successfully utilized to identify regulatory elements and the sequences within that contribute to gene expression. Arrangement of elements within a synthetic promoter can result in very precise transgene expression [42] and non-specific expression that results from the additional elements present within the "full-length" promoter sequences is avoided. The DR5 auxin promoter is a highly active synthetic promoter that contains tandem direct repeats of the auxin responsive TGTCTC element and has been used to study auxin response mechanisms in plants [43]. The CaMV35S promoter with its duplicated enhancer sequence in a synthetic context [44] has also been widely used to increase gene expression levels in numerous applications in plants. The mac promoter – a hybrid promoter constructed with part of the mas promoter and the CaMV35S enhancer region, gives even higher expression in plant tissues compared to the double enhancer CaMV35S promoter [45].

Additional element stacking or inclusion of intronic sequences into the promoter can potentially provide even greater precision in the specificity and intensity of promoter activity. As synthetic promoters are generally much smaller than full-length promoters and packed with regulatory element sequences, use of chromatin insulators [46] should be carefully assessed to minimize any undesirable interference or interaction between closely spaced synthetic promoter sequences. A more detailed analysis of synthetic promoters is presented in Section 6.

4. Enhancing introns and their role in gene transcription

Introns are intervening DNA sequences largely distributed throughout gene coding regions in eukaryotes. Introns are involved in alternative splicing and these intergenic sequences can greatly influence gene expression at different stages by affecting transcription, nuclear export, translation, and mRNA decay. Many of the polyubiquitin genes and some other highly expressed genes contain leading introns in the 5′UTR, which are often used and therefore included as part of the promoter [41,47]. These leading introns are frequently utilized to enhance transgene expression. The stimulatory effect of leading introns on gene expression is called intron-mediated enhancement (IME) and, unlike classical transcriptional enhancers, which are located upstream of the transcriptional start site, enhancing introns are located downstream and adjacent to the transcriptional start site [48]. Quantitative enhancement of transcription has been demonstrated using different introns. Some introns (*e.g. UBQ10* and *PhADF1*) can also disrupt tissue-specific patterns when they are fused to heterologous promoters and validated in *Arabidopsis* [49,50]. The intron 1 of the *OsMADS1* gene appears to contain a regulatory region essential for flower-preferential expression of this gene in rice [51]. Although the specific molecular mechanisms of IME are not fully understood, intron splicing factors can directly interact with RNA polymerase II (RNA Pol II) and, therefore, influence transcriptional levels [52]. Transcription rates and RNA Pol II phosphorylation can also influence the splicing process, indicating a tight link between these two processes. IME also involves a physical gene looping interaction between the promoter and the gene terminator *via* the 5′ and 3′ splice sites [53]. This gene looping mechanism may increase transcriptional activity as the result of the recycle of the polymerase from the terminator to the promoter for transcription re-initiation [54].

Intron-mediated transcriptional regulation can also be driven by *cis*-acting elements such as enhancers present within intron sequences. A good approach for detecting enhancer-containing introns is to evaluate introns in both orientations, with splice sites and branch points preserved. Other motif sequences present in leading introns have been implicated in quantitative and qualitative regulation of expression levels and are important IME signals. For example, the *UBQ10* intron contains numerous IME signals like the CGATT motif involved in high expression levels [55,56]. The *Arabidopsis ACT1* intron also contains specific *cis*-elements responsible for high expression levels in reproductive tissues [57]. Although the specific mechanisms of IME remain unclear, it appears that IME may influence gene expression in a number of different ways. Regardless, leading introns can have important benefits in applications requiring either high expression levels or tissue-specific gene expression.

Similar to synthetic promoters, "synthetic introns" can also be generated to appropriately evaluate the contribution of the intron or intron components to enhancement of gene expression. Synthetic introns, containing duplicated copies of the same or different intron-derived or non-intron-derived *cis*-elements, can be constructed using the same tools as used for synthetic promoters (Fig. 5). Synthetic introns could be useful for a wide array of transgenic applications, leading to exceptionally high levels of expression or the regulation of more elaborate phenotypes. Cloning of enhancers within the introns adds greater flexibility

to a synthetic promoter approach by allowing the evaluation of different combinations of enhancers and elements within a leading intron and/or the proximal promoter (Fig. 5B). Intron-derived elements from a soybean polyubiquitin promoter give increased gene expression if placed within the proximal promoter region but greater enhancements result if they are placed within the context of a synthetic intron (Hernandez-Garcia and Finer, unpublished). Each intron-derived element should be systematically evaluated in transgenic systems to determine suitability for transgene regulation. Similarly, intron-less promoters may be potentially engineered with synthetic introns to modify and improve their driven expression patterns, although special attention should be directed toward the splicing mechanisms. Disruption of the transgene splicing can certainly negatively affect IME [52].

5. Genomic and transcriptomic approaches for promoter discovery

A large toolbox of plant promoters is needed to provide more diversity and finer regulation of gene expression patterns, required for various transgenic applications including trait stacking using multi-gene stack vectors. The growing number of genome sequences for several important plants will greatly facilitate the identification of large numbers of promoters. Thirty-one promoters were isolated and partly characterized using DNA sequences identified in genomic databases of grapevine (*Vitis* spp.) [58]. In addition, we have created the Soybean Upstream Regulatory Element (SURE) database containing information on functional expression analysis of ~100 promoters (http://www.oardc.ohio-state.edu/SURE/; Finer, unpublished data), which were isolated using the soybean genome sequence and characterized using transgenic approaches.

A large number of genes in eukaryotes such as *Drosophila*, mice, and humans are not regulated by single promoters but multiple alternative promoters [59–61]. Study of alternative promoter usage has received little attention in plants, although advances in genomics and sequencing technologies would accelerate studies of alternate promoter usage in plants. The MAP kinase gene *OsBWMK1* appears to be differentially expressed as the result of alternative promoter usage, which generates two *OsBWMK1* transcript variants with different transcriptional initiation sites in rice [62]. The rice *LAGGING GROWTH AND DEVELOPMENT 1* (*LGD1*) gene encodes multiple transcripts that contain different transcription start sites [63]. Functional promoter analysis performed for two promoter variants (LGD1.1 and LGD1.5) indicates that alternative LGD1 promoters are functional but drive different expression levels [63]. Large-scale analyses of alternative first exons associated with tissue- and/or development-specific transcription have also been conducted on *Arabidopsis* and rice [64,65]; however, pairing mechanisms of alternative first exon and alternative promoter usage in plants remain elusive. A comprehensive annotation of alternative promoter usage along with expression data, such as that recently reported for *Drosophila* [59], would greatly facilitate the discovery of plant promoters with a more predictable regulation.

The pairing of genome sequence databases to genome-wide transcriptomic analysis will also be quite useful for prediction of numerous *cis*-elements with diverse functionality. For example, genome-wide *in silico* analyses using available genome sequences and transcriptomic data revealed several putative *cis*-elements within promoters from sucrose transporter genes and cold- and dehydration-responsive genes from *Arabidopsis*, rice, and soybean [66,67]. Similarly, large-scale prediction of numerous *cis*-acting elements involved in plant hormones responses was performed in *Arabidopsis* [68]. A comprehensive study of the *Arabidopsis* transcriptome using microarray analysis was also conducted to identify calcium-regulated promoters and the *cis*-elements conferring Ca^{2+} response [69]. Over a thousand putative *cis*-regulatory elements responsive to biotic and abiotic stress were identified after performing global analysis of stress gene expression data in *Arabidopsis* [1].

In silico genome-wide analyses of *cis*-elements are important to gain a better understanding of global gene regulation at the organismal level. However, the small size of motif sequences that are recognized during genome-wide analyses using current prediction algorithms, frequently leads to the identification of a tremendous number of putative elements. Moreover, the presence of DNA sequence motifs alone is not sufficient to identify functional protein binding sites, which is highly influenced by other several factors. Many copies of a short sequence motif can be present in large genomes; however, pending position and accessibility, only a small portion of those copies may be functional and enable *in vivo* protein binding [70]. Previous reports estimated that 1 in 500 motif instances are functional and bound *in vivo* in mammalian genomes. For example, The GATA-binding factor 1 (GATA1) motif is present at approximately 8 million instances, but only about 15,000 are bound by the GATA1 protein in mouse cells [71,72].

To overcome this methodological constraint, new high-throughput methodologies that integrate mapping of functional transcription binding sites along with histone modifications have been devised and successfully applied in animal genomes [70]. STARR-seq (self-transcribing active regulatory region sequencing) allowed quantification of enhancer activity of millions of candidate enhancer-containing regions in the *Drosophila* genome [73]. STARR-seq technology uses screening vectors containing small pieces of DNA isolated from sheared genomic DNA and cloned downstream of a minimal promoter – which allows potential enhancers to be self-transcribed. Reporter libraries are then transfected into candidate cells and transcript analysis of potential enhancers is assayed by high-throughput paired end sequencing. This novel approach would be a powerful tool to predict gene regulatory regions and enhancers in plants.

6. *Cis*-element validation

Regardless of the method used for promoter region and regulatory element identification, experimental validation of promoters and their putative elements should be carried out using plant transformation and transgene expression analysis. Although expression of the native gene associated with the promoter can provide some information, promoters removed from their native context and fused to new genes frequently display different expression profiles. Promoter regions are placed upstream of reporter genes and introduced into target plant cells to determine promoter functionality. The promoter regions regulating the reporter gene can be whole uninterrupted promoter sequence regions from various plant genomes or synthetic promoters, containing various promoter sequences from different sources (Fig. 4). Because of the low background expression and efficient transcription initiation, a minimal CaMV35S core promoter has been widely used for element validation. Elements can be evaluated as monomers, dimers, tetramers, and so on, either singly or in combination.

Synthetic promoters have been pivotal for basic studies of signaling and transcriptional activation. In a pioneering study, synthetic minimal promoters harboring multiple *cis*-elements (boxes W1, W2, GCC, JERE, S, Gst1, and D) responsive to pathogen attack were used to demonstrate that defense signaling is largely conserved across species at the promoter element level [42]. This same study also provided evidence for the convergence of resistance genes and wound-induced responses at the *cis*-element level [42]. Similarly, pathogen-induced expression driven by minimal promoters that contain pathogenesis-related elements (PR1),

salicylic acid responsive elements (SARE), ethylene responsive elements (ERE), and jasmonic acid responsive elements (JAR) was characterized using an *Agrobacterium*-mediated transient expression assay, which may be a potential tool for rapid high-throughput analysis of *in planta* pathogen responses [74]. Minimal promoters containing different combinations and copy number of elements (DRE/CRT, ABRE, G-box, MYB, MYC, *as1*, *rps1* site 1-like) from stress-inducible promoters were also used to study drought- and salinity-mediated responses in *Arabidopsis* [75].

The use of minimal promoters to drive genes of interest has not been fully exploited but great potential exists for using synthetic regulatory regions in the development of transgenic crops with complex agronomic traits attained using stacked gene introductions. Due to their reduced sizes and flexibility to combine *cis*-elements with different functionality, synthetic promoters can be helpful in molecular stacking where multiple genes are manipulated simultaneously and the final size of expression vectors can be an issue during transformation. The use of *cis*-elements in synthetic promoters can also be advantageous for their reduced "leaky" and off-target gene expression that often results from additional *cis*-elements in "full-length" promoter sequences.

Although minimal promoters are fused with putative *cis*-elements to determine element functionality, these synthetic promoters must be appropriately designed to generate meaningful results. Promoter inducibility and strength of expression are largely affected by spacing and motif copy number in synthetic promoters [42]. For instance, an increase in motif copy number did not necessarily result in enhanced promoter activity for some elements, and a single copy was sometimes sufficient to activate a pathogen-induced response [42]. Excess of transcription factor-binding sites due to multiple copies of the same *cis*-element can also deplete the native cognate transcription factors, reducing transcription; thus, a threshold of copy number should be determined for each element. Element spacing is also difficult to predict with minimal promoters as the size of each DNA fragment is theoretically determined by the physical conformation of the protein/*cis*-element binding complex, and should be optimized for each *cis*-element. This is an important consideration for construction of synthetic promoters combining elements with different functionality (Fig. 4) and, thus, represents a major challenge for synthetic promoter design.

7. Gene expression validation tools and promoter analysis

High-throughput recognition of promoters and putative *cis*-elements is now feasible for several important crops, as a result of the increasing number of genome sequences and the significant advances in new sequencing technologies for transcriptome analysis. However, functional characterization of isolated promoter sequences may be limited by the availability of gene introduction and expression systems. Many different validation systems exist and the data output that is generated is largely determined by the expression system used for functional analysis. Validation utilizing transient expression analysis is relatively simple and rapid while stable transformation is the most complex and lengthy, but provides the most robust information on promoter function.

A wide range of transient expression systems has been developed for rapid analysis of promoter function. Although these systems may not provide definitive information on gene expression patterns, they are powerful tools to gauge promoter function and strength. Agro-infiltration of leaves with *Agrobacterium* cultures is commonly used for rapid evaluation of transient gene expression, but this system may not be useful for spatiotemporal promoters driving expression in other organs [76]. *Agrobacterium*-mediated transient expression was used to characterize expression patterns provided by synthetic inducible promoters as a response to biotic and abiotic stresses [77]. We devised and successfully used an automated image collection and analysis system for rapid and quantitative transient expression analysis of numerous promoters using lima bean cotyledons and particle bombardment [41,78,79].

Although transient expression seems to be a good predictor of intensity of *in planta* expression, careful consideration should be exercised during the interpretation of transient expression data. Introduction of DNA into plant cells often leads to very high initial levels of gene expression compared to stable expression [80]. For direct DNA introduction, this higher expression levels is likely attributable to extrachromosomal expression and the relatively high amounts of introduced DNA that could eventually overload the regulatory machinery in a targeted cell and lead to gene silencing. Use of viral silencing suppressors with transient expression systems reduces and extends gene expression [81]. Extrachromosomal transient expression [82] is not subject to chromatin-based gene regulation and, thus, transient expression assays does not reflect the full functionality of promoters and their *cis*-acting elements.

Cell cultures and protoplasts transformed using different DNA introduction methods can also yield very rapid gene expression analysis results but tissue specificity of certain classes of spatiotemporal promoters is difficult to properly evaluate due to the obvious absence of differentiated tissues. Nevertheless, cells cultures provide some of the highest transformation efficiencies reported and are useful to analyze promoters responsive to a broad range of treatments such as light, temperature, hormones, metabolites, and pathogen derived elicitors [83].

Since recovery of large numbers of transgenic plants can require extensive time and effort, the need for rapid and simple stable transformation assays is imperative in the genomic era. Production of stably transformed hairy roots takes weeks rather than months, allowing rapid introduction and validation of numerous DNA constructs [84]. Hairy roots are induced by *Agrobacterium rhizogenes* and are transgenic, with a slightly altered phenotype from non-transformed roots. Although gene expression analysis in only root tissue can limit the usefulness of evaluation for many promoter families that drive preferential expression in aboveground plant tissues, hairy roots are generally a good preliminary indicator of expression patterns for certain promoters such as constitutive and inducible promoters [41,85]. Hairy roots are also valuable for analysis of tuber/storage organ promoters [86]. Evaluation of gene expression in hairy roots can also be useful for identification of some functional *cis*-elements (Fig. 3; Hernandez-Garcia, unpublished data). In addition, hairy roots are optimal for Green Fluorescent Protein (GFP) detection as they do not contain chlorophyll, which can impact GFP detection due to chlorophyll autofluorescence [87].

In planta characterization is obviously the best means of elucidating the function of promoters in stably transformed target species. Although *Arabidopsis* and tobacco have been extensively used as models for characterization of promoters, the absence of certain anatomical structures in these model plants may impede the study of gene expression of some types of promoters such as spatiotemporal promoters. Transgene expression driven by promoters in model plants may not be fully predictable either as some of the factors that regulate promoters may not exist in heterologous systems. For example, seed-specific promoters isolated from genes encoding storage proteins in barley (*B-Hor*, *D-Hor*) and wheat (*HMW-Glu*) showed off-target GFP expression in transgenic rice [88]. Some constitutive promoters such actin and polyubiquitin promoters have shown contrasting expression levels in different plants including *Gladiolus*, freesia, Easter lily, tobacco, rice, and rose [89,90]. The rice ubiquitin 2 (*rubi2*) promoter has shown limited expression in transgenic switchgrass [91]. Differential performance of promoters in monocot and dicot plants has also been reported. The CaMV35S promoter drives relatively low expression levels in

different monocot plants. Thus, promoters and their contributing *cis*-elements intended for biotechnological applications should be functionally characterized in the same plant species of interest.

Transformation protocols for different crops such as soybean, maize, rice and wheat are well established yet generation of transgenic lines remains time-consuming and laborious. Yet, for proper analysis of promoters in stably transformed plants, a large number of transformants should be generated and screened for gene expression, which in some cases limit the use of stably transformed plants for high-throughput analysis of promoters.

The following considerations deserve special attention for expression analysis in stably transformed tissues: (a) the size of the isolated promoter can affect gene expression as distant *cis*-acting elements can be missed during promoter cloning, (b) the absence of transcription factors required to interact with *cis*-elements can modify the expected expression patterns in heterologous systems, (c) random integration of promoter constructs in independent transformation events can result in large variability in expression due to position effect and potential loss of chromatin-mediated regulation, and (d) transgene copy number and the DNA arrangement can either increase the intensity of transgene expression or cause gene silencing.

8. Concluding remarks

To gain a correct understanding of gene function in both basic and applied research studies, introduced transgenes should be precisely regulated. With an increased number of available genes and gene stacking technologies, the possession of a well-furnished toolbox of promoters and *cis*-elements is needed to drive more complex agronomic traits such as yield enhancement and metabolic engineering. Considering that plants are comprised of several tens of thousands of genes with an exceptionally wide range of functionalities, it would not be surprising that an astonishing high number of promoters and regulatory elements remains to be discovered. This would lead to significant improvement in the regulation of numerous phenotypes and expression patterns driven by currently available promoters.

Advances in bioinformatics will lead to accelerated discovery of novel promoters and their *cis*-elements. For example, the development of more robust computational methods is still needed to precisely predict the boundaries of promoters, especially at the 5′ end. Cloning and characterization of complete “full length” promoters is preferred over truncated promoters to avoid missing potential enhancers located in the distal promoter region. In addition, computer algorithms with better prediction of *cis*-elements are required to circumvent the tremendous number of false positive elements that are detected using currently available tools. These new algorithms should be developed to efficiently recognize short semi-conserved sequences, and predict cognate transcriptional regulators.

Genome-wide transcriptomic analyses using high-throughput sequencing technologies such as RNA-seq will not only lead to a better understanding of global regulation of gene expression but also to the identification of novel promoters and *cis*-elements in economically important crops. Sharing the large amounts of transcriptomic data in public depositories may also help the selection of the best candidate promoters for specific expression patterns. The development and improvement of existing gene expression systems for simple, rapid and more predictable expression patterns will be critical to validate the large amount of expression data generated using integrated transcriptomic and bioinformatic approaches. Further development of expression validation tools is also still needed to allow high-throughput analyses of promoter activity under an assorted array of conditions.

Acknowledgments

The author’s research is supported by the United Soybean Board, the North Central Soybean Research Program, and State and Federal funds appropriated to The Ohio State University/Ohio Agricultural Research and Development Center. This work was partially supported by a fellowship from CONACYT, Mexico, to CMHG.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2013.12.007.

References

[1] C. Zou, et al., *Cis*-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*, Proceedings of the National Academy of Sciences of the United States of America 108 (2011) 14992–14997.
[2] C. Molina, E. Grotewold, Genome wide analysis of *Arabidopsis* core promoters, BMC Genomics 6 (2005) 25, http://dx.doi.org/10.1186/1471-2164-1186-1125.
[3] S.K. Burley, R.G. Roeder, Biochemistry and structural biology of transcription factor IID (TFIID), Annual Review of Biochemistry 65 (1996) 769–799.
[4] T.I. Lee, R.A. Young, Transcription of eukaryotic protein-coding genes, Annual Review of Genetics 34 (2000) 77–137.
[5] M. Bulger, M. Groudine, Functional and mechanistic diversity of distal transcription enhancers, Cell 144 (2011) 327–339.
[6] E. Calo, J. Wysocka, Modification of enhancer chromatin: what, how, and why? Molecular Cell 49 (2013) 825–837.
[7] S. Henikoff, Nucleosome destabilization in the epigenetic regulation of gene expression, Nature Reviews Genetics 9 (2008) 15–26.
[8] C. Jin, G. Felsenfeld, Nucleosome stability mediated by histone variants H3.3 and H2A.Z, Genes and Development 21 (2007) 1519–1529.
[9] R.B. Deal, S. Henikoff, Histone variants and modifications in plant gene regulation, Current Opinion in Plant Biology 14 (2011) 116–122.
[10] S. Feng, S.E. Jacobsen, W. Reik, Epigenetic reprogramming in plant and animal development, Science 330 (2010) 622–627.
[11] J.T. Odell, F. Nagy, N.H. Chua, Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter, Nature 313 (1985) 810–812.
[12] E. Lam, P.N. Benfey, P.M. Gilmartin, R.X. Fang, N.H. Chua, Site-specific mutations alter *in vitro* factor binding and change promoter expression pattern in transgenic plants, Proceedings of the National Academy of Sciences of the United States of America 86 (1989) 7890–7894.
[13] P.N. Benfey, N.H. Chua, The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants, Science 250 (1990) 959–966.
[14] S.-H. Park, et al., Analysis of the APX, PGD1 and R1G1B constitutive gene promoters in various organs over three homozygous generations of transgenic rice plants, Planta 235 (2012) 1397–1408.
[15] Y. Wang, W. Zhang, J. Cao, D. McElroy, R. Wu, Characterization of *cis*-acting elements regulating transcription from the promoter of a constitutively active rice actin gene, Molecular and Cellular Biology 12 (1992) 3399–3406.
[16] L. Tian, et al., Analysis and use of the tobacco eIF4A-10 promoter elements for transgene expression, Journal of Plant Physiology 162 (2005) 1355–1366.
[17] S.J. Gurr, P.J. Rushton, Engineering plants with increased disease resistance: how are we going to express it? Trends in Biotechnology 23 (2005) 283–290.
[18] J.F. Petolino, J.P. Davies, Designed transcriptional regulators for trait development, Plant Science 201/202 (2013) 128–136.
[19] A. Peremarti, et al., Promoter diversity in multigene transformation, Plant Molecular Biology 73 (2010) 363–378.
[20] C. Potenza, L. Aleman, C. Sengupta-Gopalan, Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation, In Vitro Cellular & Developmental Biology – Plant 40 (2004) 1–22.
[21] T. Kawakatsu, F. Takaiwa, Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains, Plant Biotechnology Journal 8 (2010) 939–953.
[22] D.J. Cosgrove, Loosening of plant cell walls by expansins, Nature 407 (2000) 321–326.
[23] K. Maeo, et al., Sugar-responsible elements in the promoter of a gene for beta-amylase of sweet potato, Plant Molecular Biology 46 (2001) 627–637.
[24] M. Zourelidou, M. de Torres-Zabala, C. Smith, M.W. Bevan, Storekeeper defines a new class of plant-specific DNA-binding proteins and is a putative regulator of patatin expression, Plant Journal 30 (2002) 489–497.
[25] N.C. Bisht, A. Jagannath, P.K. Burma, A.K. Pradhan, D. Pental, Retransformation of a male sterile barnase line with the barstar gene as an efficient alternative method to identify male sterile–restorer combinations for heterosis breeding, Plant Cell Reports 26 (2007) 727–733.
[26] M. Padidam, M. Gore, D.L. Lu, O. Smirnova, Chemical-inducible, ecdysone receptor-based gene expression system for plants, Transgenic Research 12 (2003) 101–109.

[27] A. Martinez, C. Sparks, C.A. Hart, J. Thompson, I. Jepson, Ecdysone agonist inducible transcription in transgenic tobacco plants, Plant Journal 19 (1999) 97–106.
[28] A. Okuzaki, K. Konagaya, Y. Nanasato, M. Tsuda, Y. Tabei, Estrogen-inducible GFP expression patterns in rice (*Oryza sativa* L.), Plant Cell Reports 30 (2011) 529–538.
[29] T. Aoyama, N.H. Chua, A glucocorticoid-mediated transcriptional induction system in transgenic plants, Plant Journal 11 (1997) 605–612.
[30] H.A. Roslan, et al., Characterization of the ethanol-inducible alc gene-expression system in *Arabidopsis thaliana*, Plant Journal 28 (2001) 225–235.
[31] A. Maizel, D. Weigel, Temporally and spatially controlled induction of gene expression in *Arabidopsis thaliana*, Plant Journal 38 (2004) 164–171.
[32] S.W. Bang, et al., Characterization of the stress-inducible OsNCED3 promoter in different transgenic rice organs and over three homozygous generations, Planta 237 (2013) 211–224.
[33] N. Yi, et al., Analysis of the Wsi18, a stress-inducible promoter that is active in the whole grain of transgenic rice, Transgenic Research 20 (2011) 153–163.
[34] A. Pellegrineschi, et al., Stress-induced expression in wheat of the *Arabidopsis thaliana* DREB1A gene delays water stress symptoms under greenhouse conditions, Genome 47 (2004) 493–500.
[35] K. Yamaguchi-Shinozaki, K. Shinozaki, A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress, Plant Cell 6 (1994) 251–264.
[36] C. Jiang, B. Iu, J. Singh, Requirement of a CCGAC *cis*-acting element for cold induction of the *BN115* gene from winter *Brassica napus*, Plant Molecular Biology 30 (1996) 679–684.
[37] K. Yamaguchi-Shinozaki, K. Shinozaki, Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses, Annual Review of Plant Biology 57 (2006) 781–803.
[38] N. Kovalchuk, et al., Defensin promoters as potential tools for engineering disease resistance in cereal grains, Plant Biotechnology Journal 8 (2010) 47–64.
[39] S.H. Hwang, I.A. Lee, S.W. Yie, D.J. Hwang, Identification of an OsPR10a promoter region responsive to salicylic acid, Planta 227 (2008) 1141–1150.
[40] A. Himmelbach, et al., Promoters of the barley germin-like GER4 gene cluster enable strong transgene expression in response to pathogen attack, Plant Cell 22 (2010) 937–952.
[41] C. Hernandez-Garcia, et al., High level transgenic expression of soybean (*Glycine max*) GmERF and Gmubi gene promoters isolated by a novel promoter analysis pipeline, BMC Plant Biology 10 (2010) 237, http://dx.doi.org/10.1186/1471-2229-1110-1237.
[42] P.J. Rushton, A. Reinstadler, V. Lipka, B. Lippok, I.E. Somssich, Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling, Plant Cell 14 (2002) 749–762.
[43] T. Ulmasov, J. Murfett, G. Hagen, T.J. Guilfoyle, Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements, Plant Cell 9 (1997) 1963–1971.
[44] F. Guerineau, A. Lucy, P. Mullineaux, Effect of two consensus sequences preceding the translation initiator codon on gene expression in plant protoplasts, Plant Molecular Biology 18 (1992) 815–818.
[45] L. Comai, P. Moran, D. Maslyar, Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements, Plant Molecular Biology 15 (1990) 373–381.
[46] J.E. Phillips-Cremins, V.G. Corces, Chromatin insulators: linking genome organization to cellular function, Molecular Cell 50 (2013) 461–474.
[47] C.M. Hernandez-Garcia, A.P. Martinelli, R.A. Bouchard, J.J. Finer, A soybean (*Glycine max*) polyubiquitin promoter gives strong constitutive expression in transgenic soybean, Plant Cell Reports 28 (2009) 837–849.
[48] H. Le Hir, A. Nott, M.J. Moore, How introns influence and enhance eukaryotic gene expression, Trends in Biochemical Sciences 28 (2003) 215–220.
[49] S. Emami, D. Arumainayagam, I. Korf, A.B. Rose, The effects of a stimulating intron on the expression of heterologous genes in *Arabidopsis thaliana*, Plant Biotechnology Journal (2013), http://dx.doi.org/10.1111/pbi.12043.
[50] Y.M. Jeong, J.H. Mun, H. Kim, S.Y. Lee, S.G. Kim, An upstream region in the first intron of petunia actin-depolymerizing factor 1 affects tissue-specific expression in transgenic *Arabidopsis* (*Arabidopsis thaliana*), Plant Journal 50 (2007) 230–239.
[51] J.S. Jeon, S. Lee, G. An, Intragenic control of expression of a rice MADS box gene OsMADS1, Molecules and Cells 26 (2008) 474–480.
[52] N. Skoko, M. Baralle, S. Tisminetzky, E. Buratti, InTRONs in biotech, Molecular Biotechnology 48 (2011) 290–297.
[53] A.M. Moabbi, N. Agarwal, B. El Kaderi, A. Ansari, Role for gene looping in intron-mediated enhancement of transcription, Proceedings of the National Academy of Sciences of the United States of America 109 (2012) 8505–8510.
[54] S. Lykke-Andersen, C.K. Mapendano, T.H. Jensen, An ending is a new beginning: transcription termination supports re-initiation, Cell Cycle 10 (2011) 863–865.
[55] G. Parra, K. Bradnam, A.B. Rose, I. Korf, Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants, Nucleic Acids Research 39 (2011) 5328–5337.
[56] A.B. Rose, T. Elfersi, G. Parra, I. Korf, Promoter-proximal introns in *Arabidopsis thaliana* are enriched in dispersed signals that elevate gene expression, Plant Cell 20 (2008) 543–551.
[57] A. Vitale, R.J. Wu, Z. Cheng, R.B. Meagher, Multiple conserved 5′ elements are required for high-level pollen expression of the *Arabidopsis* reproductive actin ACT1, Plant Molecular Biology 52 (2003) 1135–1151.
[58] Z.T. Li, K.H. Kim, J.R. Jasinski, M.R. Creech, D.J. Gray, Large-scale characterization of promoters from grapevine (*Vitis* spp.) using quantitative anthocyanin and GUS assay systems, Plant Science 196 (2012) 132–142.
[59] P. Batut, A. Dobin, C. Plessy, P. Carninci, T.R. Gingeras, High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression, Genome Research 23 (2013) 169–180.
[60] A.R. Banday, S. Azim, M. Tabish, Alternative promoter usage and differential expression of multiple transcripts of mouse Prkar1a gene, Molecular and Cellular Biochemistry 357 (2011) 263–274.
[61] K. Miyazaki, et al., Differential usage of alternate promoters of the human stress response gene ATF3 in stress response and cancer cells, Nucleic Acids Research 37 (2009) 1438–1451.
[62] S. Koo, et al., Identification and characterization of alternative promoters of the rice MAP kinase gene OsBWMK1, Molecules and Cells 27 (2009) 467–473.
[63] S. Thangasamy, P.-W. Chen, M.-H. Lai, J. Chen, G.-Y. Jauh, Rice LGD1 containing RNA binding activity affects growth and development through alternative promoters, Plant Journal 71 (2012) 288–302.
[64] W.-H. Chen, G. Lv, C. Lv, C. Zeng, S. Hu, Systematic analysis of alternative first exons in plant genomes, BMC Plant Biology 7 (2007) 55, http://dx.doi.org/10.1186/1471-2229-1187-1155.
[65] N. Kitagawa, et al., Computational analysis suggests that alternative first exons are involved in tissue-specific transcription in rice (*Oryza sativa*), Bioinformatics 21 (2005) 1758–1763.
[66] O. Ibraheem, C.E. Botha, G. Bradley, In silico analysis of *cis*-acting regulatory elements in 5′ regulatory regions of sucrose transporter gene families in rice (*Oryza sativa* Japonica) and *Arabidopsis thaliana*, Computational Biology and Chemistry 34 (2010) 268–283.
[67] K. Maruyama, et al., Identification of *cis*-acting promoter elements in cold- and dehydration-induced transcriptional pathways in *Arabidopsis*, rice, and soybean, DNA Research 19 (2012) 37–49.
[68] Y. Yamamoto, et al., Prediction of transcriptional regulatory elements for plant hormone responses based on microarray data, BMC Plant Biology 11 (2011) 39, http://dx.doi.org/10.1186/1471-2229-1111-1139.
[69] H.J. Whalley, et al., Transcriptomic analysis reveals calcium regulation of specific promoter motifs in *Arabidopsis*, Plant Cell 23 (2011) 4079–4095.
[70] R.C. Hardison, J. Taylor, Genomic approaches towards finding *cis*-regulatory modules in animals, Nature Reviews Genetics 13 (2012) 469–483.
[71] Y. Zhang, et al., Primary sequence and epigenetic determinants of *in vivo* occupancy of genomic DNA by GATA1, Nucleic Acids Research 37 (2009) 7024–7038.
[72] Y. Cheng, et al., Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression, Genome Research 19 (2009) 2172–2184.
[73] C.D. Arnold, et al., Genome-wide quantitative enhancer activity maps identified by STARR-seq, Science 339 (2013) 1074–1077.
[74] W. Liu, et al., Bacterial pathogen phytosensing in transgenic tobacco and *Arabidopsis* plants, Plant Biotechnology Journal 11 (2013) 43–52.
[75] L. Hou, et al., Construction of stress responsive synthetic promoters and analysis of their activity in transgenic *Arabidopsis thaliana*, Plant Molecular Biology Reporter 30 (2012) 1496–1506.
[76] M. Santos-Rosa, A. Poutaraud, D. Merdinoglu, P. Mestre, Development of a transient expression system in grapevine via agro-infiltration, Plant Cell Reports 27 (2008) 1053–1063.
[77] W. Liu, M. Mazarei, M. Rudis, M. Fethe, C.N. Stewart, Rapid *in vivo* analysis of synthetic promoters for plant pathogen phytosensing, BMC Biotechnology 11 (2011) 108, http://dx.doi.org/10.1186/1472-6750-1111-1108.
[78] J.M. Chiera, et al., Isolation of two highly active soybean (*Glycine max* (L.) Merr.) promoters and their characterization using a new automated image collection and analysis system, Plant Cell Reports 26 (2007) 1501–1509.
[79] C.M. Hernandez-Garcia, J.M. Chiera, J.J. Finer, Robotics and dynamic image analysis for studies of gene expression in plant tissues, Journal of Visualized Experiments 39 (2010), http://dx.doi.org/10.3791/1733 http://www.jove.com/index/details.stp?id=1733
[80] I. Potrykus, Gene transfer to plants: assessment and perspectives, Physiologia Plantarum 79 (1990) 125–134.
[81] T. Dhillon, J.M. Chiera, J.A. Lindbo, J.J. Finer, Quantitative evaluation of six different viral suppressors of silencing using image analysis of transient GFP expression, Plant Cell Reports 28 (2009) 639–647.
[82] L. Rossi, B. Hohn, B. Tinland, Integration of complete transferred DNA units is dependent on the activity of virulence E2 protein of *Agrobacterium tumefaciens*, Proceedings of the National Academy of Sciences of the United States of America 93 (1996) 126–130.
[83] J. Sheen, Signal transduction in maize and *Arabidopsis* mesophyll protoplasts, Plant Physiology 127 (2001) 1466–1475.
[84] A. Makhzoum, P. Sharma, M. Bernards, J. Trémouillaux-Guiller, Hairy roots: an ideal platform for transgenic plant production and other promising applications, in: D.R. Gang (Ed.), Phytochemicals, Plant Growth, and the Environment, Springer, New York, 2013, pp. 95–142.
[85] E.H. Hughes, S.-B. Hong, J.V. Shanks, K.-Y. San, S.I. Gibson, Characterization of an inducible promoter system in *Catharanthus roseus* hairy roots, Biotechnology Progress 18 (2002) 1183–1186.
[86] V. Veena, C. Taylor, *Agrobacterium rhizogenes*: recent developments and promising applications, In Vitro Cellular & Developmental Biology – Plant 43 (2007) 383–403.
[87] C. Wu, J.M. Chiera, P.P. Ling, J.J. Finer, Isoxaflutole treatment leads to reversible tissue bleaching and allows for more effective detection of GFP in transgenic

soybean tissues, In Vitro Cellular & Developmental Biology – Plant 44 (2008) 540–547.
[88] A. Furtado, R.J. Henry, F. Takaiwa, Comparison of promoters in transgenic rice, Plant Biotechnology Journal 6 (2008) 679–693.
[89] Y.H. Joung, K. Kamo, Expression of a polyubiquitin promoter isolated from *Gladiolus*, Plant Cell Reports 25 (2006) 1081–1088.
[90] K. Kamo, A. Blowers, F. Smith, J.V. Eck, R. Lawson, Stable transformation of *Gladiolus* using suspension cells and callus, Journal of the American Society for Horticultural Science 120 (1995) 347–352.
[91] M.N. Somleva, et al., Production of polyhydroxybutyrate in switchgrass, a value-added co-product in an important lignocellulosic biomass crop, Plant Biotechnology Journal 6 (2008) 663–678.