

Cis-regulatory elements used to control gene expression in plants

Róża Bilas¹ · Katarzyna Szafran^{1,2} · Katarzyna Hnatuszko-Konka¹ ·
Andrzej K. Kononowicz¹

Bla Bla

Received: 1 February 2016 / Accepted: 2 August 2016 / Published online: 10 August 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Plant biotechnology is a dynamically developing science, which comprises many fields of knowledge. Novel plant genetic engineering findings highly influence the improvement of industrial production. These findings mostly concern *cis*-regulatory elements, which are sequences controlling gene expression at all developmental stages. They comprise of promoters, enhancers, insulators and silencers, which are used to construct synthetic expression cassettes. Examples of most important *cis*-regulatory elements are reviewed in the present paper. Variability among core promoters content and distal promoter regions impedes evaluation of interactions between them during the artificial promoters construction. Synthetic promoters and artificial expression cassettes trigger a significant increase in gene expression level, better properties and quality of a product. Accumulating knowledge about gene promoters, *cis* sequences and their cooperating factors allows uniform expression systems and highly predictable results.

Keywords Promoter · Enhancer · Silencer · Insulator · Plant

Introduction

In the past century, biotechnology became one of the most vigorously evolving branches of science due to its main purpose—supporting numerous economic sectors. Improvement of specific organisms is all-important especially in the case of plants since their unassisted adaptation is time-consuming and frequently fails. The wide implementation of novel methods depends on the economic benefits and public opinion. So far, plants have been modified to produce vaccines, hormones, enzymes and antibodies (Shaaltiel et al. 2007; Rabindran et al. 2009; López et al. 2010; Sirko et al. 2011; Ma and Wang 2012; Madanala et al. 2015).

Essential research objectives include development of innovative, advanced and easily applicable methods to obtain plants of high productivity (Bock 2013) under both normal and stress conditions (Sharma et al. 2011; Gouiaa et al. 2012). Frequently applied transformation methods include gene gun and agroinfection with *Agrobacterium tumefaciens* (Ma and Wang 2012). Product accumulation in a particular organ and especially in a target cellular compartment is highly preferred and commonly practiced (Potenza et al. 2004).

Recombinant protein production in plants offers many advantages. Production and storage costs are relatively low and the product quality is high. Post-translational protein modification pathways are very similar to mammalian ones. Furthermore, infection of the product by mammalian pathogens is avoided (Tekoah et al. 2015). Unlike mammalian cells or microorganisms, use of plants allows a substantial increase in production scale due to product accumulation in given localization, primarily in the ER (López et al. 2010; Martinez et al. 2011), as well as by enlarging the culture volume or acreage. However, systems based on plants are not without disadvantages since they require longer production

Róża Bilas and Katarzyna Szafran have contributed equally to this work and are co-first authors.

✉ Katarzyna Hnatuszko-Konka
kath@biol.uni.lodz.pl

¹ Faculty of Biology and Environmental Protection,
Department of Genetics, Plant Molecular Biology and
Biotechnology, University of Lodz, Banacha Street 12/16,
90-237 Lodz, Poland

² Present address: Faculty of Biology, Department of Applied
Microbiology, University of Warsaw, Miecznikowa Street 1,
02-096 Warsaw, Poland

time (depending on the development cycle length) and arouse a lot of social controversies. The latter relates mainly to the possibility of cross-bordered pollution caused by uncontrolled transgene spread onto wild organisms after genetically modified crops are introduced into the environment. In this case, chloroplast transformations (Daniell 2006; Ma and Wang 2012) or industrial-scale production under laboratory conditions may be a reasonable solution (Martinez et al. 2011; Miao et al. 2008).

Undoubtedly, extending the use of plants in biotechnology depends on simultaneous genetic engineering development, which provides the necessary tools for species modifications in order to meet the expected demands (Bock 2013). The essential role in controlling processes is not played by encoding sequences, as was previously thought, but by regulatory elements (Venter and Botha 2010). Multiplicity and diversity of regulatory sequences make their exact composition and specification time-consuming. It has to be emphasized that regulation occurs at each stage of gene expression, so that the control mediated by regulatory sequences is a comprehensive action. Strictly controlled implementation of genetic information allows obtaining a final product of required parameters.

Detailed examination of the effects caused by a given regulatory sequence is vital as they present various activities from enhancing to silencing or insulating. This knowledge is required to strictly plan research concerning expression levels and specificity. Preliminarily experimental construction of different expression cassette combinations is highly recommended until the composition is optimized to guarantee best results (Mehrotra et al. 2011; Liu et al. 2013; Twyman et al. 2013).

Due to zygote divisions, each cell of a plant organism possesses an identical set of genetic information encoded in its DNA. However, synthesis of the entire set of products encoded in a genome would be metabolically unfavorable. Moreover, it could also disrupt vital functions. Hence, during differentiation, cells develop features characteristic of the particular tissue type by transiently losing the ability to transcribe genes encoding products irrelevant for a given tissue. However, this loss is not permanent so dedifferentiation to the primary state is possible (Kaufmann et al. 2010).

Independently of differentiation, genes that encode key substrates for primary vital functions, called housekeeping genes, are always active. One method to identify housekeeping genes in the genome is gene knockout (Allen 2008), which involves removal of unknown genes and analysis of the cells vitality afterwards.

Gene expression can be constant or can show temporal or spatial regulation pattern. Summarizing the crucial stages, final product quality and amount are established by transcription initiation at the promoter region, post-transcriptional and post-translational modifications, and depend on

further protein transport and accumulation efficiency (Li et al. 2012).

It is the commonly known fact, that gene expression is spatially compartmented and multistage in eukaryotic organisms. Each stage is precisely determined and controlled by specific factors. Availability of template DNA and presence of genes in Euchromatin region are prerequisites for information implementation. Primary regulation occurs during functional initiation complex assembling on the DNA strand. Next, RNA polymerase II localizes an initiation site (IS) and attaches to it. Both stages require participation of transcription factors (TF). After elongation and termination, immature, unstable pre-mRNA is formed. The subsequent maturation step comprises splicing (excising of introns and joining of exons) and attachment of 5' cap and 3' polyA tail to the transcript ends. Mature mRNA is then transported to the cytoplasm where translation takes place (Klug and Cummings 2003; Phillips 2008). Translation is based on the formation of mRNA-ribosome subunits complexes. It allows nucleotide sequence translation into linear amino acids sequence, which results in the formation of the proteins primary structure. Due to further protein modifications, higher structures are obtained, which non-protein components are able to bind to. A final functional form appears as a result of these modifications (Kawaguchi and Bailey-Serres 2002; Phillips 2008; Bock 2013; Browning and Bailey-Serres 2015). Translation initiation efficiency depends on the nucleotides flanking start codon ATG, called TIC (translation initiation context) and on its distance from transcript 5' end (Koul et al. 2012). As Koul et al. (2012) reported, TAAACAATGG is the most effective TIC arrangement among dicotyledonous plants.

Types of regulatory elements

Many elements present in plant genome control the gene expression level through interactions with DNA or regulatory proteins at every stage of implementation of genetic information. Regulators are classified in terms of their structure as *cis* sequences and *trans* factors. *Cis* regulatory sequences are linear nucleotide fragments of non-coding DNA. Their localization and orientation in relation to genes and activity is various (Venter and Botha 2010). Plant regulatory sequences are located directly in the transcribed DNA strand: promoters, enhancers, silencers and insulators; or may be added during post-transcriptional modifications: 5'cap, poly-A tail, signal sequences (Vaughn et al. 2012). Specific regulatory proteins, called *trans* elements, interact with *cis* sequences and other proteins to form active complexes. Organization of all eukaryotic genomes is similar and most of regulatory elements are universal. However, substantial differences occur among elements assigned

to particular tissues e.g. tissue/organ-specific promoters (Twyman 2003; Venter and Botha 2010).

This paper reviews *cis* sequences that are present directly in the DNA or may be attached during post-transcriptional and post-translational modifications.

Structure of eukaryotic gene promoters

Gene promoters, located upstream of the gene coding sequence, enable initiation of transcription by the presence of RNA polymerase II binding sites. This enzyme is attached during the sequential binding of specific proteins, called transcription factors, which results in transcription initiation (Russell 1996; Porto et al. 2014).

The structure of promoters, including plant promoters, can be divided into two regions—core promoter and distal region comprising enhancers and silencers (Fig. 1). Both these regions contain *cis* elements, which proteins, known as *trans* factors may bind to (Klug and Cummings 2003; Peremarti et al. 2010; Porto et al. 2014).

Core promoter

The core promoter regulates the appropriate initiation of transcription by RNA polymerase II (Juven-Gershon and Kadonaga 2010). It may be composed of TATA box, initiator region (Inr), CAAT box and downstream promoter element (DPE). However, not all these regions are always present at each promoter. Occurrence of some promoter elements, such as TFIIB recognition element (BRE) and motif ten element (MTE) was not confirmed in plants (Juven-Gershon and Kadonaga 2010; Porto et al. 2014). Some promoter classifications are based on the presence or absence of the TATA box and other elements.

The TATA box, located 25–30 bp upstream of the transcription start site (TSS), is usually flanked with GC-rich

regions (Lewin 2001). It is the first observed and described conservative DNA fragment, approximately 8 bp in length, composed of AT base pairs with the consensus sequence TATA(A/T)A(A/T) (Twyman 2003). The transcription factors and RNA polymerase II bind to the TATA box in a specific order. The transcription factor TFIID is bound first and it forms a complex with TFIIA. TFIIB attaches to this complex via TATA-binding protein (TBP), which is a part of TFIID, and as a result of direct interaction with the DNA strand. RNA polymerase II is bound in the next step and TFIIF attaches to it. Thereupon TFIIE and TFIIH join the complex and the full transcription apparatus is formed (Kwak and Lis 2013).

The initiator region (Inr) is the most common element of the core promoter (Juven-Gershon and Kadonaga 2010). However, both promoters lacking this region, and promoters without neither Inr nor TATA box have been described. In this case, the initiation of transcription may occur in a number of places and is not strictly determined. Inr sequence usually covers up with TSS (Juven-Gershon and Kadonaga 2010) or is at least close by it (Twyman 2003). Inr has the consensus sequence YYCARR (Twyman 2003). The remaining promoter elements are located at a precise distance from the initiator adenine. Therefore, the space between this adenine (referred to as A₊₁) and a definite sequence is used to specify the position of this sequence (Juven-Gershon and Kadonaga 2010). In the presence of the TATA box, the initiator region cooperates with it upon the transcription initiation. Since Inr is recognized by TFIID, it may also replace the TATA box (Kadonaga 2012).

The following, frequently observed conservative promoter element is the CAAT box, located approximately 80 bp upstream of TSS. It is able to act in both directions and significantly influence the expression efficiency. CAAT is highly susceptible to mutations hence it is considered as the most important factor affecting transcription effectiveness. In plants, a similar AGGA box has been identified in

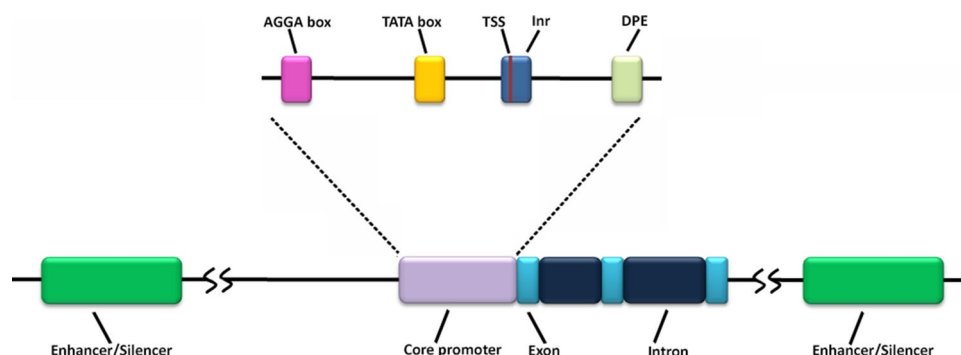


Fig. 1 Example of a plant gene organization. Not all of the depicted elements are universal. The core promoter may consist of the AGGA box, the TATA box, Inr and DPE. TSS is the transcription start site, Inr

is the initiator and DPE is the downstream promoter element. Figure based on Klug and Cummings (2003)

place of the CAAT box (Roa-Rodríguez 2003; Porto et al. 2014).

In the promoters lacking the TATA box, DPE (downstream promoter element) may be responsible for binding TFIID protein. This element is usually located 28–32 bp downstream of adenine A₊₁ of the Inr. Occasionally TATA, Inr and DPE coexist. Even in the case of a single nucleotide alteration between the DPE and Inr, TFIID cannot bind properly to the promoter and the expression efficiency decreases significantly. In plants, several copies of DPE are present upstream the transcription start site and regulate the response to external stimuli (Kutach and Kadonaga 2000; Sawant et al. 2001).

Some core promoter elements are not typical for plants. One of them is BRE, a sequence which may increase or decrease transcription efficiency by binding TFIIB. Two types of this sequence are known: BREu is located directly upstream of the TATA box, while BREd is placed directly downstream of the TATA box (Deng and Roberts 2005; Juven-Gershon and Kadonaga 2010). The GC-rich sequence, is not present in all plants, but plays an important role in many animal promoters, such as binding transcription factor Sp1, which boosts expression efficiency. It may act in both directions, be located at various distances from the TSS and be present in several copies (Lewin 2001; Roa-Rodríguez 2003). Moreover, in the animal gene promoters, motif ten element was identified, which may be found directly upstream of the 5' end of DPE and may affect TFIID recognition. MTE may function independently from DPE, but requires presence of Inr (Juven-Gershon and Kadonaga 2010).

Numerous core promoter elements identified and characterized previously are not frequently observed. GARE and GT1 motifs, thymine- and cytosine-rich repeats, I and L sequences and pyrimidine-rich sequence inside 5'-UTR may be the examples of uncommon elements. GARE motif, located between 139 and 145 bp downstream of the TSS, takes part in the gibberellin response. The G1 motif (ATGGTGGTTGG), which may be found 168–178 bp downstream of the TSS is one of the elements responsible for the light response. TC-rich repeats which facilitate the defense against stressors are located 17–26 bp downstream of the TSS and sequence TTTCTCTCTCTC in 5'UTR region ensures high-level protein expression (Porto et al. 2014).

Distal promoter regions

Unlike previously described elements, enhancers, and silencers are located further from the TSS. They may be located several thousands of base pairs away from the TSS and be included in non-coding intron sequences (Peremarti et al. 2010). Enhancers are more common than silencers.

Enhancers and other distal *cis* sequences will be described thoroughly in the following sections.

A capability of acting under specific environmental conditions or in the definite tissues is an important feature of distal promoter elements. Diverse availability of transcription factors in various tissues provides this opportunity. Thereupon some promoters are constitutive, while the others may be tissue-specific or inducible (Peremarti et al. 2010). Identification of enhancers acting in both directions or under certain conditions is exceptionally important for plant biotechnology, as it may facilitate construction of valuable artificial promoters.

Types of gene promoters

The promoters utilized in plant biotechnology vary in the terms of efficiency, site, and period of action. They may be active throughout all developmental stages in each tissue (constitutive promoters) or in particular tissues or development-stage (tissue-specific promoters). Some promoters may require specific stimuli for activation (inducible promoters) or be operating in the specific developmental stage. It has to be emphasized that promoters isolated from dicotyledonous plants are more efficient while used in dicots. The same rule applies for promoters isolated from monocots (Park et al. 2010).

Constitutive promoters

Odell et al. (1985) analyzed the sequence of constitutive promoter *CaMV 35S*, isolated from *Cauliflower mosaic virus*. They showed that, if used intact, it was active in all tissues of many plant species. The analysis of promoters formed of various *CaMV 35S* subdomain combinations revealed miscellaneous activity in distinct tissues (Benfey et al. 1990). It proves the specific character of several subdomains (Peremarti et al. 2010). *CaMV 35S* is more frequently used in transformation of dicots, as it is more efficient in those plants as compared to monocots. It may be a result of distinct transcription factors availability in the foregoing groups of plants (Urwin et al. 1997; McCarter 2009). Expression of genes controlled by the *CaMV 35S* promoter may be disturbed when the plant is infected with nematodes (Peremarti et al. 2010). In spite of the mainly constitutive character, *CaMV 35S* activity may be lower in some tissues or organs, such as generative organs (Porto et al. 2014).

CaMV 35S is not the unique viral promoter used for plant genetic modifications. Some of them, such as *Sugarcane bacilliform virus* promoter (*ScBV*) and *Commelina yellow mottle virus* promoter (*CoYMV*) are effective both in monocots and dicots. The latter was carefully analyzed on the example of monocots, but it was shown that provides

constitutive expression also in potato (Medberry et al. 1992; Torbert et al. 1998).

Moreover, many efficient constitutive promoters have been isolated from plants. They are considered safer than viral promoters (Potenza et al. 2004). The most widely applied constitutive promoter utilized for monocots genetic modification (*ZmUbi1*) has been isolated from corn. Natively, it regulates the production of ubiquitin (Cornejo et al. 1993). Several other ubiquitin promoters, derived from *Arabidopsis*, potato, sunflower or rice, are known but not frequently applied (Sharma and Sharma 2009). Ubiquitin is a protein present in all eukaryotic organisms. It takes part in various processes such as DNA repair, protein degradation, proper chromatin folding, and heat stress response (Ciechanover 1998). *ZmUbi1* promoter ensures high expression levels, particularly in young leaves and roots. Concurrently with plant maturation, production of the protein of interest decreases (Christensen et al. 1992; Park et al. 2010; Shepherd et al. 2014). Inducing heat-stress on the plant may increase *ZmUbi1* efficiency. It was proved by the incubation of plover at temperatures 7 and 12 °C higher than applied previously. *ZmUbi1* promoter may be used as a fusion with intron of ubiquitin-1 gene to obtain higher expression level in flowers and seeds (Stoger et al. 1999). The described promoter is ten-fold more active in monocots than the *CaMV* 35S promoter (Christensen et al. 1992). Polyubiquitin promoters *PvUbi1* and *PvUbi2*, isolated from switchgrass, analyzed by Mann et al. (2011), demonstrated particularly high constitutive expression. They are more efficient than *ZmUbi1*, *CaMV* 35S and *OsAct1* (actin promoter derived from rice). However, this is a relatively new discovery, thus polyubiquitin promoters are not widely used currently.

Gene promoters responsible for the synthesis of proteins present in the entire plant are a potential subject of research in the field of constitutive promoters. However, some exceptions to this rule have been observed, for example, in *Arabidopsis thaliana* and *Oryza sativa* plants. Due to the function of actin (essential protein forming the cytoskeleton), actin promoters such as *OsAct1* are constitutive, (McElroy et al. 1990) but yet, their activity may take a different form when natural promoters are cloned into expression constructs. For example, the *Act2* promoter, which is a modified *Arabidopsis* actin promoter, is substantially constitutive, but does not promote protein expression in hypocotyl, seed coat and microsporangia of *Arabidopsis*. Furthermore, the modified rice actin promoter is not active in the xylem of this plant (Zhang et al. 1991; An et al. 1996).

To allow the constitutive expression of two transgenes, a bidirectional promoter construct can be used. Zhang et al. (2008) obtained a bidirectional promoter by assembling two unidirectional promoters, fused in opposite directions. For example, the diverted core promoter *CaMV* 35S was assembled to the 5' end of the intact *CaMV* 35S promoter.

CaMV 35S is a unidirectional promoter, while the resulting construct provided the strong constitutive expression of reporter genes, flanking the promoter in both directions. What is interesting, higher expression level of one reporter gene was associated with lower expression of the other gene. Both promoter fragments require the same transcription factors and were probably competing for a certain amount of those proteins, which led to the described phenomenon (Zhang et al. 2008).

Tissue-specific promoters

Numerous tissue-specific promoters, providing expression in the certain type of tissue, have been characterized so far. The remaining tissues do not contain the heterologous protein, which is a significant feature. Exogenous protein production may lead to various disorders and developmental abnormalities such as dwarfism or programmed cell death. As the heterologous peptide is present in the exact organ, hazard for the plant metabolism is much lower, than while the constitutive promoters are used (Sharma and Sharma 2009).

Promoters active in the seeds are one of the most important specific promoters. The seeds of legumes, rice, wheat and corn are the most substantial sources of food all around the world. Increasing their nutritive value is significant, particularly for the poorer countries' citizens. Actually, seeds are abundant in proteins, so they may be transformed to the essential source of heterologous protein with the specific promoters. In monocots, storage substances are accumulated in the endosperm, while in dicots cotyledons are used for this purpose (Chen et al. 2007). Hence, endosperm-specific promoters are exceptionally beneficial. For example, pro-insulin, laccase, β -carotene and *Infectious bursal disease virus* have been produced in seeds up until now (Sharma and Sharma 2009).

Promoters' specificity may vary with the host organism species. Alike constitutive promoters, those obtained from monocots may not be equally active in dicots (and *vice versa*). Occasionally, application of promoter considered as specific may result in non-specific heterologous gene expression (Park et al. 2010). For example, wheat transformation with glutenin promoter *LMWG1D1* led to the efficient expression within the endosperm. Glutenin is a low molecular weight protein present natively in wheat seeds. However, when the same promoter was used for tobacco transformation, the expression level in seeds was between 4 and 20-fold lower, and the heterologous protein was detected in leaves. For rice modifications, *Gt1* promoter specific for endosperm may be used (Stoger et al. 1999).

Protein profile of dicots seeds may be modified successfully with various promoters such as *zE19* derived from corn. This tandem promoter region of zein gene provides

efficient expression in the cotyledons, however, the presence of reporter proteins in alternative organs may be observed. Depending on the species, its localization and concentration will vary (Chen et al. 2007).

β -Conglycinin promoter from soy, and α -globulin promoter from cotton are further seed-specific promoters with the peak activity during the middle and late phase of seed maturation (Chamberland et al. 1992; Sunilkumar et al. 2002). β -phaseolin promoter from bean is another promoter of described activity (Bustos et al. 1989). β -1,3-glucanase promoter from peas (*PsGNS2*) is an interesting example of seed-specific construct, as it provides transgene expression in the seed coat. This feature may be useful for labelling genetically modified seeds with colors to facilitate their distinction from wild type seeds (Buchner et al. 2002).

Another group of widely used gene promoters are those that regulate gene expression exclusively in fruits. They are used to either increase the nutrition value of these organs or to produce edible vaccines (Potenza et al. 2004). *E8* promoter, isolated from tomato, unlike other promoters regulating fruit maturation genes, provides given type of expression. Nonetheless, a slight production of reporter protein in the anthers has been reported. *E8* is used for antigen production in fruits and improving the typical tomato aroma and flavour (Deikman and Fischer 1988). Further important application of fruit-specific promoters is related to maintaining their freshness after harvest (Hernandez-Garcia and Finer 2014).

Targeting the transgene expression to anthers is essential for establishing male sterility, which is an important feature facilitating the control over prevalence of genetically modified varieties. This goal may be achieved by applying tobacco *TA29* or *Arabidopsis A9* promoters (Koltunow et al. 1990; Bisht et al. 2004). The chimeric form of those two regulatory elements may be used as well (Bisht et al. 2007). Mariani et al. (1990) obtained male-sterile rapeseed by introducing the bacterial cytotoxic ribonuclease (barnase) into the anthers. Interestingly, when the transgenic females were crossed with males expressing the barnase inhibitor, the progeny plants were fertile.

Promoters of genes controlling starch and glycoprotein storage in the tubers of potato, cassava, sweet potato, taro or yam are widely used to improve their nutritional value or fight viral and fungal infections (Chakraborty et al. 2000; Mohammed et al. 2000). Several regulatory elements, such as dioscorin promoter *pDJ3S*, patatin class I promoters (*B33* and *PAT1*), sporamin promoter and β -amylase promoters are tuber-specific. *PAT1* and *B33* are most active in the vascular tissues during initial tuber development and later on in parenchyma. As they contain sucrose-inducible sequences, the expression efficiency may be boosted by supplementation with this sugar (Jefferson et al. 1990; Liu et al. 1990; Hernandez-Garcia and Finer 2014).

Time-specific promoters

Most of promoters which activity is not observed throughout ontogenesis, may be induced by the external stimuli. Promoters specific to both the certain tissue and developmental stage, as described previously, are another type of time-specific promoters.

Artificial external inducers of biotechnologically useful promoters should meet several requirements such as notable specificity, rapid mechanism of action and easy application. Moreover, the expression level ought to be dose-dependent and translocation of the compound shall be dynamic to allow the utilization in a distance from the site of action. From an economic point of view, such substances should be cheap and widespread. The above criteria seem to be complex, however many inducible protein expression systems have been described previously. Inducers such as plant hormones, ethanol and heat shock are applied (Potenza et al. 2004; Corrado and Karali 2009; Sharma and Sharma 2009).

Various types of stressors are most frequently used for inducible promoter activation. Many stress-reacting promoters contain at least two common *cis* elements. Thus, a response to different stressors is possible by introducing the same promoter. The ACGTG(G/T)C sequence, called ABA-responsive element (ABRE) is one of the mentioned *cis* elements (Bonetta and McCourt 1998). Another sequence, known as the dehydration-responsive element (TACCGA-CAT), regulates the defense reaction to low temperatures, drought and dangerous salination (Yamaguchi-Shinozaki and Shinozaki 1994; Hernandez-Garcia and Finer 2014).

The majority of inducible systems consist of two elements. The first controls the production of transcription factor, which binds to another element—the target promoter. For example, ethanol-inducible promoter isolated from *Aspergillus nidulans* comprises *alcR* gene and *alcA* promoter (Caddick et al. 1998). *AlcR* codes transcription factor ALCR, which may be activated by the presence of ethanol or acetaldehyde. Functional ALCR binds to the promoter of the *alcA* gene, coding alcohol dehydrogenase. This system applies to plant modifications as follows: *alcR* is controlled by constitutive promoter, usually *CaMV 35S*, while *alcA* promotes the transgene expression. It ensures high sensitivity and efficiency. Expression starts immediately after inducer application and is dose-dependent (Salter et al. 1998; Roslan et al. 2001; Klose et al. 2013).

Moreover, heterologous gene transcription may be activated with phytohormones, such as abscisic acid (ABA), auxins, and gibberellins, of which, the lattermost react rapidly. Genes being indirectly controlled by the gibberellic acid (GA_3) presence are provided with silencers typically bound with DELLA proteins. This interaction prevents transcription by restraining attachment of transcription factors. The presence of GA_3 in the environment leads to DELLA

proteolysis and thereby induces expression of genes remaining under control of this promoter (Dill et al. 2004; Hauvermale et al. 2012; Davière and Achard 2013).

To obtain infection-resistant plant varieties, promoters induced by wounds and pathogen infections are utilized. Mechanical damage by pathogens provides two types of reaction: rapid (necrosis of cells surrounding the wound) and systemic (Corrado and Karali 2009). It was proved, that jasmonic acid (JA) and, sometimes, ethylene, take part in the wound signalling pathway. Upon injury, a signal protein called systemin, increases jasmonic acid synthesis, and the formed JA induces production of the defensive proteins (Stratmann 2003; Delano-Frier et al. 2013).

Further, well-described wound-inducible regulatory element is the promoter of proteinase inhibitor II (*pin2*). It demonstrates slight constitutive activity, which may be significantly increased in the epidermis by the injury (Keil et al. 1990; Corrado and Karali 2009). These types of promoters frequently comprise many *cis* elements with just a few responsible for the described specificity. By isolating these elements and joining them into synthetic promoters, efficient transgene expression induced by injury may be obtained (Rushton et al. 2002).

Synthetic promoters

The structure of plant promoters allows to construct various sets of regulatory elements. Multiplication of the elements, their reorganization and ligation with synthetic and native *cis* sequences are used to obtain synthetic promoters (Bhullar et al. 2003; Mehrotra et al. 2011; Venter and Botha 2010). An optimal plant synthetic promoter should have precisely defined specificity, immediate inducibility, the versatility of applications and should guarantee the best efficiency of transformation (Venter and Botha 2010).

Two main methods are applied during synthetic promoter composition and analyses. They include modification of distal promoter comprising *cis* regulatory blocks and creation of two-gene set including one gene encoding *trans* factor and another gene, being influenced by this *trans* factor (Venter and Botha 2010). In latter method, expression of the gene encoding *trans* factors supplies activators to a multiplied block of *cis* elements in the other gene. Therefore, induced, increased and stable expression of the introduced gene is guaranteed. *Trans* factor production is controlled by the strong constitutive or specific promoter.

There are various applications of synthetic promoters. For example Liu et al. (2011) integrated *cis* elements induced by phytohormones, produced during pathogen infection, with a minimal promoter *CaMV* 35S. The obtained regulatory cassette was applied to control a gene of red fluorescence protein *pporRFP* (*Porites porites* red fluorescent protein) in rice. Rapid induction of gene expression after biotic stress

application was observed, thus enabling early and easy localization of infected regions in plants *in vivo*. Recent studies focus also plant on synthetic promoters containing UPT boxes (upregulated by transcription activator-like effectors) responsible for resistance to various plant pathogens. Moreover, pathogen responsive and pathogen inducible (PRPI) promoters are used to obtain production and storage of a given protein in tissues exposed to pathogen (Mehrotra et al. 2011).

The bidirectionality of plant promoters is a phenomenon noticed in nature (e.g. oleosin promoter) and applied into studies (Zhang et al. 2008). It allows to obtain increased and more stable expression of two genes in multigene constructs. This ability is used mainly to study plant disease resistance and gene silencing (Zheng et al. 2011). Synthetic plant bidirectional promoters could be obtained by ligation of an opposite oriented promoter with the 5' end of the other promoter (Zhang et al. 2008; Venter and Botha 2010). It was proved that the orientation of core promoter elements is essential for bidirectionalization and should be consistent with the direction of a given gene (Zheng et al. 2011). The space between promoters can be left unused or supplemented with *cis* regulators boosting the activity of both promoters at the same time thereby resulting in effective enhance of the whole cassette expression (Venter and Botha 2010). Zheng et al. (2011) constructed a bidirectional promoter module (*mPtDrl02*) containing a methyl jasmonate-inducible *PtDrl02* promoter from poplar and 35S minimal promoter linked to its 5' end, in opposite orientation. Final bidirectional promoter was tested within two constructs: GFP/*mPtDrl02*/GUS and GUS/*mPtDrl02*/GFP. Studies confirmed that basal expression depended on gene direction relatively to the promoter and forward direction was more efficient. Methyl jasmonate-induced *cis*-elements, in both directions, provided GUS and GFP expression comparable to expression obtained with unidirectional promoters.

Plants have many polygenic traits, which causes some difficulties in gene modifications and transfers (Gouiaa et al. 2012; Bock 2013). Plant bidirectional promoters and gene stacking can be used to avoid gene silencing in multigene cassettes and decrease the amount of transferred DNA (Shamloo-Dashtpajardi et al. 2015). Kumar et al. (2015) constructed a cassette combining a bidirectional promoter obtained from *ZmUbi1* promoter and bicistronic gene organization. They introduced this construct to corn and obtained efficient production of four genes. Expression efficiency was higher than when *ZmUbi1* was used to express one of the used genes.

Cis sequences

Cis elements directly influence gene regulation. Their sequence is usually not constant so they may be found merely through hypothetical localization and analysis of

interacting proteins (Sharma et al. 2011). Depending on the type, *cis* sequences are present in different copy numbers, as well as variable distances and orientation in relation to the gene (Venter and Botha 2010). There are two main types of *cis* sequences—enhancers and silencers. Current analyses reveal this basic division in a new light. It is difficult to unequivocally classify some of the *cis* sequences as enhancers or silencers since they may have some additional properties or both listed functions at the same time. Therefore, *cis* sequences are more frequently differentiated on the basis of their activity in a given model system. *Cis* sequences that recur in genes ensuring identical functions are classified as motifs e.g. sequences specific for genes induced by phytohormones, abiotic stress or specific developmental stages. Advancements in discovering and researching *cis* sequences and cooperating factors lead to uniform applications and highly predictable results in further studies.

Cis-regulatory sequences research

At present, there are two main approaches for *cis* regulatory research focused on the elements localization, orientation and cooperation. This leads to findings of their more efficient novel applications. The aim is to obtain methods independent from gene coding sequences and mostly based on promoter organization because some conservative motifs have lost their previous function and some specific elements appeared among genomes during evolution (Shamloo-Dashtpajardi et al. 2015). One approach, widely presented in this review, is in vitro practical verification of elements chosen by predictions based on former experiences. Described method often uses DNA–protein interactions. Nowadays, many important transcription factors are well-known and characterized. It led to the development of methods for *cis* regulatory elements recognition e.g. DNase I footprinting (Allen 2008), electrophoretic mobility shift assays, ChIP microarrays, surface plasmon resonance, and yeast one-hybrid systems (Dare et al. 2008). The second main approach uses in silico algorithms and databases to calculate the probability of predicted regulatory element rightness (Mehrotra et al. 2011; Agarwal et al. 2014). This method also allows collection of various sets of predicted elements at the same time and their classification and verification by a number of statistical tests to reduce the amount of elements to the most probable (Sharma et al. 2011). However, the last identification step also requires biological confirmation (Jung et al. 2011). Recently, Shamloo-Dashtpajardi et al. (2015) described the whole process of in silico research. They primarily aimed to define the exact occurrence of *cis* regulatory elements, which may be accepted statistically in research, and determine promoters' features and differences among expression levels of genes similar in function. Dare et al. (2008) studied co-regulated genes activated by one transcription factor,

PAP1. They observed decreased gene activation after inserting mutations into DNA fragments appointed to comprise *cis* regulatory elements.

Enhancer sequences

Plant enhancers are located at different, often considerable, distances upstream or downstream of the promoter sequence. They enhance gene expression through cooperation with specific transcription factors (Mehrotra et al. 2011). Few models of their activity have been published. One theory suggests that a specific protein is bound to the enhancer sequence. Thereby, a loop is formed by the connection of enhancer and transcription apparatus. Transcription efficiency improvement is considered to be a result of the described conformational alteration (Alberts et al. 2002). In most cases, interaction with the core promoter does not depend on the distance between the enhancer and promoter. This property corresponds to the DNA looping. Recent studies reported a correlation between enhancers and chromatin state, including dependence on histone modifications. Enhancers increase the availability of DNA to replication and transcription factors, as well as intensify these processes (Barrett et al. 2012). The effect of interactions between enhancers and promoters depends on the distance between them only if it is as significant as several thousand base pairs and then inhibits the enhancers' activity. Both weak plant enhancers like *Nos*, and strong enhancers such as *supP* or *CaMV 35S* are known and used (Yang et al. 2011). Correlation between number of enhancer copies and gene expression level was analyzed. Initially, constructs containing several copies of enhancers showed additive effect. However, as the number of copies increased—the effect was more often synergistic, or even general decreased gene activity was observed (Liu and Bao 2009).

Liu and Bao (2009) constructed a synthetic basal *ocs* promoter using an enhancer derived from octopine synthase gene *ocs*. The objectives were to test the influence of a number of enhancer inverted repeats (IR) and their distance from the TATA-box on the expression of the *gus* reporter gene in tobacco leaves. It was shown that the level of enhancement depended on the distance from the TATA-box and decreased over its lengthening (Liu and Bao 2009). The results also confirmed that the increase in the number of copies to three enhanced gene activity. Each additional copy weakened the total effect. This indicates that each given pair of enhancer and promoter has an optimal number of enhancer copies that should be defined during preliminary analysis preceding essential study. Analysis should include testing marker gene expression using constructs containing different numbers of enhancer copies in relation to the chosen promoter.

Promoter activation by an adjacent enhancer could be non-specific, therefore allowing formation of chimeras that

have novel traits, while maintaining their original properties. Investigation of interactions between stigma-specific promoter *SLG* from *Brassica* with *LAT52* enhancer, typical of pollen, and *AGL5* enhancer specific for carpel were held. It was shown that the promoter performed novel activity in the tissues specific for the given enhancer without any influence to its primary specificity (Yang et al. 2011).

Combining enhancers derived from constitutively active genes with specific promoters results in various effects. The *CaMV 35S* enhancer can increase activity in tissue-specific promoters to constitutive levels in tissues in which they are not originally active, or may be active occasionally (Yang et al. 2011). Zheng et al. (2007) analyzed three variants of constructs, each containing *CaMV 35S* promoter in enhancer function for: vascular-specific promoter *AAP2* with *gus* gene, ovary-specific promoter *AGL5* with *iaaM* gene, promoter typical of early embryogenesis and sexual reproduction *PAB5* with barnase gene. The presence of the promoter acting as an enhancer triggered constitutive activity of all genes with different expression levels. However, the use of enhancers derived from constitutive promoters in constructs containing specific promoters may result in numerous dimorphisms and decrease the transformation efficiency, by activating expression in both generative and vegetative tissues (Yang et al. 2011).

Synthetic gene expression enhancing cassettes

The need of creating various enhancing element combinations in distal promoter regions in order to increase expression efficiency and a risk of ambiguous effects was perceived. It led to attempts to construct common and properly working set of *cis* elements. Numerous analyses resulted in the construction of *Pcec* enhancing cassette (complete expression cassette). It consists of two main components: *Pmec* (minimum expression cassette including core promoter with TATA-box) and a block of variable *cis* elements (Venter and Botha 2010). To obtain *cis* blocks, only enhancers of highly expressed genes are chosen, especially those acting with promoters induced by phytohormones, biotic and abiotic stress or environmental conditions (Koul et al. 2012). Synthetic cassettes can be also used to investigate plant promoters bidirectionalization e.g. bidirectional promoter obtained from two *Pmec* cassettes regulated by plant *cis* elements (induced with phytohormones, stress or environmental conditions) was successfully tested in tobacco (Mehrotra et al. 2011). Koul et al. (2012) tested seven cassettes comprising various promoters and different numbers and types of enhancers in tomato cultures (*Lycopersicon esculentum*). Two distinct methods were used to evaluate the results. One was based on the expression of *uidA* glucuronidase gene, and the other on the activity of *cryIAc* gene encoding Δ -endotoxin—an insecticide originally derived

from *Bacillus thuringiensis*. Cassettes consisted of elements identified in the most active genes and *PcamIII* containing doubled enhancer sequence from *CaMV 35S* promoter were examined. As it was shown that they are exceptionally effective in regulation of *uidA*, they were used in the construct comprising *cryIAc*. *Pcec* application resulted in the highest (0.8%) amount of endotoxins in total soluble protein (TSP). *Pcam* cassette proved to be less effective (0.13% endotoxins in TPS). Further research provided by Mehrotra et al. (2011) focused on cassettes containing plant *cis*-motifs ACGT and GT involved in protection against plant pathogens. Depending on the different combinations and distances from TATA box they obtained between two to sevenfold higher expression.

Studies on constructing the synthetic cassettes of enhancers show how laborious this task is. However, at the same time it has huge potential, helping to obtain high expression levels.

Insulator sequences

Insulators are a group of genetically different elements, which vary amongst species. However, they induce a common effect. During plant transformation, high heterologous gene expression in a given tissue is required as well as constitutive expression of reporter genes, which are used to verify the modification effectiveness. *Cis*-elements proximity characterized by different activity levels interrupts prospective interpretation of genetic information (Singer et al. 2012). Consequently, insulators application in multigene constructs is ergonomically beneficial. Insulator sequences guarantee genes autonomy because they determine the compartments in the molecular environment of DNA strands, whereby separating the influence of chromatin state and regulators related to adjacent genes (Papadakis et al. 2004; Yang et al. 2011).

Insulators located between the enhancer and promoter separates the enhancing effect, and disturb interactions. In some cases, a similar effect could be obtained by inserting an additional nucleotide space between the enhancer and promoter. However, nucleotide space is not equal to an insulator, which is characterized by defined composition of nucleotides. Increasing the enhancer-promoter distance in DNA can block the enhancer influence. In fact, the efficient transfer of such large components, precise determination of their length, and composition are highly complicated (Yang et al. 2011). Characterization of the insulator group responsible for enhancer-promoter interactions is crucial for their profitable application. Enhancers and silencers can be located both upstream and downstream of the regulated gene. Consequently, insulators allow selective modulation of enhancers' undesirable influence depending on their location.

Insulators activity still needs experimental elucidation. However, it is hypothesized that they cause DNA strand looping or attach protein factors, which results in conformational changes and physical access barrier (Yang et al. 2011).

Barrier insulators may be found near the gene borders. They neutralize the position effect, and the influence of adjacent sequences on gene expression (Papadakis et al. 2004; Allen 2008), thus protecting genes against undesirable silencing by heterochromatin. Euchromatin domains are separated from heterochromatin by MAR sequences (matrix associated region), known also as SAR (scaffold attachment region), characterized by a high content of AT base pairs. MAR mediates structure anchoring in the nuclear matrix, described as loop domain model (Allen 2008). Hereby, an active replication and transcriptional domain available for *trans* factors is formed. Barrier insulators may also locally affect histone acetylation and methylation, and indirectly lead to chromatin activation (Allen 2008; Singer et al. 2012).

Some of the barrier insulators e.g. TBS sequence (transformation booster sequence) from *Petunia hybrida* genome can also act as insulators of enhancer-promoter interactions. Two types of constructs were used to investigate this activity: one included sequences providing high expression (35S enhancer, 35S promoter), the other—tissue-specific sequences (promoter *AGIP* specific for verticillatum leaves). Both constructs included the *gus* gene as well. After TBS introduction between enhancers and promoters, an obstruction of enhancing activity was reported. At the same time, it was confirmed that insulator activity depended on nucleotide content instead of sequence length. Further examined fragment, which was two times longer, had no blocking properties (Hily et al. 2009). Furthermore, it was found that besides MAR-like regions TBS element comprised fragments of unknown nature, which presumably undertake insulation functions even more than MAR. Similar research reported fourfold higher effectiveness of using *EXOB* sequence derived from λ phage as an insulator compared to TBS. Since *EXOB* is almost two times shorter than TBS, the presumption that the insulation effect is only a result of increasing the distance between enhancer and promoter was excluded (Yang et al. 2011).

Allen (2008) presented diverse examples of MAR sequences. ARS-1 provided 12-fold higher expression of GUS in tobacco suspension cultures when compared to constructs without MAR sequences, while RB7 increased expression level 60-fold. Allen (2008) also noticed that the *Agrobacterium* transformation system might hinder higher strength of MAR effect on gene expression.

Gene construct introduced during plant transformation is generally incorporated into plant DNA at a random location, which is difficult to predict. Therefore, insertion of an insulator to the construct prevents influence of both native

genome enhancers and enhancers included in the construct, especially in the case of multigene constructs. Modification efficiency increases then and variability amongst plants subjected to the process decreases.

5'UTR sequence

The 5'UTR sequence (untranslated region), called the leader sequence, is a fragment of mRNA transcript located on its 5' end. It is encoded in the DNA strand and submits to transcription but not translation, however, it regulates the latter process (Barrett et al. 2012). The sequence of 5'UTR comprises 5' cap, upstream open reading frame (uORF), guanine-rich fragment and IRES (internal ribosome entry site). Some plant UTRs contain a pyrimidine-rich fragment (5'UTR Py-rich stretch) responsible for high transcription levels (Barrett et al. 2012).

Each of these 5'UTR elements plays a specific role affecting the mRNA processing. However, the overall character of 5'UTR is important. Since the content and arrangement of nucleotides influences secondary loop structures forming, it is significant. Fragments with high GC content are considered to be strong translation inhibitors because they provoke highly stable hairpin loop formation and thus impede finding the start codon (Barrett et al. 2012).

Two types of 5'UTR fragments are distinguished. One, present in transcripts derived from housekeeping genes, is relatively short and its secondary structure is simple because of the low GC content. It does not contain additional AUG codons so translation initiation takes place in a simple manner and maintains a stable level. The other type of 5'UTRs are longer and present in developmental genes. Unlike the first type, the latter presents a more complex secondary structure caused by the higher content of GC-rich regions (Barrett et al. 2012).

Presence of 5'UTR stabilizes the transcript during transport from nucleus to cytoplasm where it protects it from endonuclease activity. Furthermore, correct transcript binding to the ribosomal subunits significantly depends on the overall nature of 5'UTR regions. Due to economic considerations, forming of active translation complex only takes place in the presence of all translation factors (Tyurin et al. 2016).

The 5' cap is a single 7-methylguanosine nucleotide followed by a number of various methylated riboses. It provides a connection between protein translational factors, small 40S ribosomal subunit and the transcript (Bock 2013). Some of the transcripts are not equipped with a 5' cap, hence the pre-initiation complex forms on the IRES sequence.

An open reading frame (ORF) is a fragment of the transcript, beginning with the initiation codon and ending with the termination codon, which undergoes translation. Some plant 5'UTRs with additional single or multiple initiation

and termination sites upstream the main AUG codon are known (Guerrero-González et al. 2014; Zhou et al. 2010). A fragment comprising of an additional frame and followed by a short sequence terminated with an ORF is called an upstream open reading frame (uORF), while the correct primary ORF is called the major or main ORF (mORF) (Hayden and Jorgensen 2007). In general, the occurrence of uORFs reduces the translation efficiency.

Ribosomal activity disruption results from the incorrect recognition of a uORF as an ORF and the onset of erroneous translation. The loss of the protein specific function may be caused by additional amino acids introduced during the described process. In some cases, the ribosome may start translation from the correct ORF after the first round of translation, by reinitiation. As it was proven by Zhou et al. (2010) with the use of mutants, the reinitiation process is strongly enhanced by the eukaryotic initiation factor 3 (eIF3) and the ribosomal protein RPL24.

Otherwise, the small subunit ignores the false start to avoid mistakes and keeps checking the strand to find the mORF. This process is called *leaky scanning* (Ivanov et al. 2010; Nyikó et al. 2009).

Nowadays, there are three properties of the uORF suggested to influence the regulation: the uORF start codon composition, the uORF length and, in consequence, the distance between the uORF and the mORF (Hayden and Jorgensen 2007).

The effect induced by the uORF sometimes can be also a consequence of the amino acid composition of the nascent protein produced by the sequence-dependent regulatory uORFs (Ebina et al. 2015). Induction of the point mutations within the AUG codon eliminates the uORF originally located in the transcript. After application of the induced mutations, an increase in translation efficiency was observed (Barrett et al. 2012). Introduction of the uORF into the construct destabilizes it by disconnecting the 5' cap, or by ribosome subunits disintegration after termination of incorrect transcript (Tran et al. 2008). In consequence, reinitiation is blocked and translation efficiency significantly decreases.

About one-third of plant transcripts possess a uORF (Guerrero-González et al. 2014; Sagor et al. 2016) and most of them can be found in genes of two groups: S-adenosylmethionine decarboxylases (AdoMetDCs) and S basic region leucine zipper (bZIP). Although it was previously thought that most of plant uORFs act in a sequence-independent manner, the well-known uORFs in transcripts of *A. thaliana AdoMetDC1* genes are a sequence-dependent type and there is constant new data suggesting such converse manner (Ebina et al. 2015).

Genes containing uORFs were found to be involved in several metabolism and signalling pathways such as polyamine, amino acids and sugar metabolism. What is important, some metabolic products of these pathways can also

act as modulators of translation (Zhou et al. 2010) and thus can play a similar role to the nascent protein in a sequence-dependent uORF regulation type. For example polyamine, a product of the polyamine metabolic pathway, can be a repressor of translation connected with the *A. thaliana AdoMetDC1* uORFs or a stimulator overcoming the presence of the uORF (Ivanov et al. 2010). Also, metabolic pathways of sugars, especially sucrose, are effectors for the bZIP uORFs (Shekhawat and Ganapathi 2014). Summarizing, the uORF controls the mORF and products of the mORF can regulate the uORF. Thus, there is an evident regulation occurring between the uORF and the mORF (Ebina et al. 2015; Hayden and Jorgensen 2007).

Computational analyses enable recognition of the positions of the uORFs in the conserved plant designated by the start and stop codons and the composition of the conserved amino acid among monocots and dicots. The listed properties do not always occur at the same time (Hayden and Jorgensen 2007).

Several studies confirmed an important role of the uORFs which should be considered in genetic engineering. Nowadays, genes from the bZIP group are readily used in transformation constructs as the encoded transcription factors often enhance the expression of the genes of interest. However, it should be remembered that many of them possess uORFs which can cause the opposite effect. For example, transgenic bananas overexpressing a gene containing a uORF—*MusabZIP53*, related to sucrose homeostasis—present abnormal growth caused by Sucrose Induced Repression of Translation (SIRT) (Shekhawat and Ganapathi 2014). In addition to uORF mutagenesis mentioned earlier, there is a novel solution to avoid the negative effect of the uORF. Recently, tobacco plants with 3–4 times higher sugar content in leaves were obtained by Thalor et al. (2012). Also Sagor et al. (2016) used ZIP genes lacking the SIRT uORFs in constructs and obtained 1, fivefold higher sugar content in tomato plants than in non-transgenic plants.

Guerrero-González et al. (2014) confirmed the negative effect of the *AtPAO3* uORF on *GUS* expression, which can be overcome by exogenous polyamines supplementation. A model proposed and tested on *A. thaliana* by Hanson et al. (2008) confirms that the constitutive expression of the bZIP11, known to be repressed by sucrose through the uORF, inhibits the production of proline and stimulates the production of phenylalanine.

Plant 5'UTRs can be combined with various promoters to obtain synthetic upstream regulatory modules (URM). They may be more effective than viral sequences like AMV or Ω , used so far. Agarwal et al. (2014) reported that 5'UTRs, originated from highly expressed genes PHOTO and GGR from *A. thaliana* enhanced *gus* gene expression in cotton and tobacco (up to 100 and 20–40 fold, respectively), while combined with *CaMV* 35S promoter.

Research led by Gouiaa et al. (2012) presents one of the solutions for multiple genes pyramiding. IRES sequence originating from tobacco NtHSF-1 gene was successfully used as a spacer and an internal ribosome attachment site to obtain higher salt tolerance by combining two well-known wheat genes encoding vacuolar ion transporters. Many studies proved that the independent over-expression of both genes leads to tolerance increase. It was a significant prerequisite to research bicistronic construct consisted of *TNHS1:IRES:TVPI* and *CaMV 35S* promoter. Results analysis showed that TVP1 expression in a co-expression system was up to two times greater than the wild type, and TVP1 over-expression in a single gene construct. Multi-level genomic, transcriptomic and cellular analysis showed stable salt and drought tolerance improvement in transgenic tobacco plants. It is worth mentioning that over-expression of H^+ -pyrophosphatase also resulted in better growth of transgenic plants when compared to wild type, under normal conditions.

To the date, numerous IRES from both viral and Eukaryotic origin were described. As Jung et al. (2011) shown in their study, sequence origin should be considered in respect of the target organism. It was confirmed by negative results obtained by using animal virus-derived IRES in constructs expressed in *O. sativa*.

3'UTR sequence and poly-A

The 3'UTR sequence is located in a transcript between the stop codon and polyadenylation signal. Since the 3'UTR is added after transcription, it is considered an important regulator of the next gene expression stage—translation. 3'UTRs act as stabilizers, enhancers and silencers. Li et al. (2012) compared 3'UTRs derived from seed storage protein (SSP) transcripts with *Nos* terminator in respect of the protein translation level and accumulation in *O. sativa* seeds. The level of mRNA transcripts was more than two times greater after the use of construct containing SSP 3'UTR than with *Nos*. Presented results confirm 3'UTR impact on enhancing transcription. It was also demonstrated that the degree of translation intensification differed and was constant for each 3'UTR whichever promoter was applied (in this case *GluC*, *Ubi-1* or *CaMV 35S*). Adenine rich elements (AREs) frequently exist within 3'UTRs. They bind proteins involved in transcript degradation, contributing to decreasing translation efficiency or strengthening the process according to macro-environmental factors or conditions, depending on the location in the cell (Yang et al. 2011).

The presence of 3'UTR and its content significantly affect the final gene expression (Papadakis et al. 2004). This property was confirmed during analysis of β -carotene content. Mentioned compound is an industrially important precursor of vitamin A. Studies on inbred maize lines reported

differences in the level of β -carotene storage. A team of researchers headed by Vignesh identified the cause of this dissimilarity (2013). Firstly, the *crtRB1* sequencing from 11 various maize lines was conducted. The *crtRB1* gene is the main gene responsible for β -carotene accumulation. The analysis revealed differences within 3'UTRs which confirmed previous assumptions, regarding the influence of sequences on variety. Ten positions of single nucleotide polymorphism (SNP) and six positions of insertion/deletion (InDe) were detected. Depending on the type of SNP, sequences responsible for high and low accumulation of β -carotene were selected. There are two hypotheses explaining this phenomenon. One assumes the influence of 3'UTR on the transcription termination stage and on the translation stabilization through incorporation of specific factors. The other considers 3'UTR cooperation with 5'UTR during loop formation, which facilitates the translation initiation (Vignesh et al. 2013).

The 3'UTR contains sequences complementary to microRNA (miRNA). Pairing of complementary sequences between mRNA and miRNA prevents translation by decreasing the amount of mRNA available in the cytoplasm. Among the 3'UTR sequences, occurrence of miRNA binding sites is more probable in longer elements. Absolute nucleotide sequence homology between miRNA and mRNA is not always required for binding to take place. In consequence longer 3'UTRs lead to decline in translation productivity (Barrett et al. 2012).

The poly-A tail is a polynucleotide sequence composed of around 250 adenines. Both poly-A and 3'UTR are added to the transcript during transcription termination. They cooperate in stabilization of mRNA structure, affect the transcript transport and its fate in the cytoplasm (Li et al. 2012). As has already been mentioned, these sequences are targets for various small RNA (snRNA) types and thus determine the number of transcripts, and indirectly regulate expression efficiency (Barrett et al. 2012).

Signal sequences: NES, NLS and KDEL

Signal sequences composed of a few dozen up to several dozen amino acids mark and direct mature proteins to their final localization in cell compartments and organelles. Choosing the target site of heterologous protein should be preceded by preliminary studies because the accumulation rate differs amongst compartments (Ma and Wang 2012). Proteins released from ribosomes to the cytoplasm must undergo post-translational processing in environment with low proteolytic potential inside the endoplasmic reticulum (ER). After chaperone-mediated folding, assembling into complexes and other modifications, proteins become completely active and ready to fulfil their functions in a cell (Takaiwa et al. 2009). They only need to be transported

to target compartments. Proteins are divided in terms of translocation type. The following types are known: proteins connected with the rough ER during the forming process, proteins released by co-translocation and proteins released freely to the cytoplasm. Further locations of free proteins in the cytoplasm are determined by signal sequences directing them to the mitochondria, peroxisomes, chloroplasts, ER or to the nucleus (Ma and Wang 2012). Signal sequences are characterized by length, hydrophobicity and overall ionic charge.

Large proteins cannot translocate spontaneously through nuclear envelope pores and their transport requires the presence of signal sequences. Nuclear localization signal (NLS) is composed of basic amino acids and directs proteins through the nuclear envelope pores into the nucleus. On the contrary, nuclear export signal (NES) marks proteins for transport out of the nucleus. Research carried out by Huang et al. (2014) led to NLS and NES identification in *A. thaliana* resistance protein RPW8.2. This discovery indicates their significance in plant resistance to a wide range of diseases.

The tetrapeptide sequence KDEL (Lys-Asp-Glu-Leu) is localized at the C-terminal protein end. KDEL signal directs proteins to the ER, where many proteins are accumulated in complexes, especially during the seeds maturation. KDEL presence provides protein protection from unfavorable cytoplasm environment inter alia proteases and oxidation or reduction processes (López et al. 2010). Recombinant protein labelling with KDEL upon production allows to precisely determine the protein localization, ensures stability and facilitates purification (Ma and Wang 2012). Martinez et al. (2011) examined the influence of KDEL on the level of the *Dengue virus* envelope protein (DV-E) accumulation in the cell suspension cultures of *Nicotiana tabacum* and *Morinda citrifolia*. Studies focused on two expression cassettes variants, both under the control of *CaMV 35S* promoter. Constructs included glycine-rich signal peptide, which directs proteins to the secretion pathway and further to the apoplast, and DV-E gene with or without KDEL sequence. The presence of KDEL caused 122% increase in the protein storage in *N. tabacum* and 110% increase in *M. citrifolia* in comparison to the variant without the additional tetrapeptide.

Eukaryotic genomes are structurally and functionally similar, which facilitates composing of expression constructs using genes derived from various species. However, studies on prokaryotic gene expression in plant cells often require specific modifications, e.g. production of phytase—enzyme from *Aspergillus niger* which catalyzes the recovery of phosphorus from phytates stored in seeds. To reduce the amount of inorganic phosphorus escaping into the environment in excess, animal feed is enriched with phytase-containing plants (Peng et al. 2006). Peng's research team (2006) analyzed the efficiency of phytase production after

an optimization of *MPHY2* gene codons for *Brassica napus* and optional addition of KDEL. It was reported that the presence of KDEL enhances both gene expression level, represented by increase in corresponding mRNA level, and protein accumulation. As a result of codon optimization, heterologous protein concentration increased and amounted from 35 to 77%.

Scientific literature gives multiple examples of using the KDEL tetrapeptide to obtain higher heterologous protein accumulation in transgenic plants. For instance, KDEL was applied to 14D9 mouse antibodies production in cell suspension cultures of *N. tabacum*. The level of accumulation increased fivefold in comparison to the wild type (López et al. 2010). Similar studies on KDEL were focused on the 7Crp protein, a Japanese cedar pollen allergen. Aggregates of 7Crp and prolamins are formed in *O. sativa* seeds in the presence of cysteine. The aggregates are then accumulated in protein bodies type I (PBs-I) which are not available for digestive enzymes. This property enables obtaining edible vaccines transferred to the lymphatic system (Takaiwa et al. 2009).

Perspectives *Cis*-regulatory sequences research

The progress in discovering *cis* sequences have significantly extended the state of knowledge and brought a lot of practical benefits in economical production planning. Many aspects of gene expression regulation remain unexplained and require detailed analysis. The number of *cis* sequences, their diversity and the awareness that not all of them have been discovered trigger further research (examples of the investigated promoters and *cis*-acting elements are listed in Table 1) (Liu and Bao 2009). Previous modifications of gene expression, especially using enhancers and promoters are now implemented in industry, which by investing in production improvement, contributes to expansion of interest in regulators. Industrial production exploits both cell suspension cultures and whole plant organisms as bioreactors.

Construction of ready to use enhancing cassettes is a novel strategy in genetic engineering. The creation process is a laborious task because it requires examination of each stage of gene expression and also physical and biochemical cell features. For example, the amount of available transcription factors (TF) is a limiting factor for transcription. Therefore, using the constructs with multiple motif copies requires sufficient quantity of TFs in the nucleus (Liu and Bao 2009). Similar problem should be considered during synthetic promoter construction. The major aim of analyses is to obtain or find stable, universal and efficient cassettes. Extensive genomes sequencing identifies successive sequences with regulatory properties, which need to be precisely characterized.

Table 1 Characteristics of the promoters and *cis*-acting elements

Sequence/type	Origin	Function	Application/reference
Promoters			
<i>CaMV 35S</i>	<i>Cauliflower mosaic virus</i>	Constitutive in dicots	Herbicide-resistant Bt corn (Saxena et al. 2004)
<i>ScBV</i>	<i>Sugarcane bacilliform virus</i>	Constitutive in monocots and dicots	Biodiesel and bioethanol production in monocots and dicots (Borysyuk et al. 2013)
<i>ZmUbi1</i>	<i>Zea mays</i> Ubiquitin gene	Constitutive in monocots	Expression of herbicide tolerance genes in corn (Lira et al. 2011)
<i>PvUbi1</i> and <i>PvUbi2</i>	<i>Panicum virgatum</i> Polyubiquitin chains genes	Constitutive	Mann et al. (2011)
<i>OsAct1</i>	<i>Oryza sativa</i> Actin gene	Constitutive	Modification of fructan biosynthesis in forage grasses (Spangenberg et al. 2010)
<i>Act2</i>	<i>Arabidopsis thaliana</i> Actin gene (modified)	Constitutive (except hypocotyl, seed coat and microsporangia)	Phytochelatin expression to obtain metal-resistant plants for phytoremediation (Meagher and Li 2002)
<i>LMWG1D1</i>	<i>Triticum aestivum</i> Glutenin gene	Endosperm-specific	Stoger et al. (1999)
<i>Gt1</i>	<i>Oryza sativa</i> Glutelin gene	Endosperm-specific	Transferrin production in rice seeds (Zhang et al. 2010)
<i>zE19</i>	<i>Zea mays</i> Zein gene	Cotyledon-specific	Chen et al. (2007)
β -Conglycinin promoter	<i>Glycine max</i>	Seed-specific, peak of activity in the middle and late phase of seed maturation	Cyanovirin-N production in soybean seeds (O'Keefe et al. 2015)
α -Globulin promoter	Cotton	Seed-specific, peak of activity in the middle and late phase of seed maturation	Reduction of gossypol content in cottonseed (Sunilkumar et al. 2006)
<i>PsGNS2</i>	<i>Pisum sativum</i> β -1,3-glucanase gene	Seed coat-specific	Labelling GM seeds (Buchner et al. 2002)
<i>E8</i>	Different tomato species	Fruit-specific, ethylene-inducible	Production of cholera toxin B in <i>Lycopersicon esculentum</i> (He et al. 2008)
<i>TA29</i>	<i>Tobacco</i>	Anthems-specific	Male sterility (Koltunow et al. 1990; Paul et al. 1992)
<i>A9</i>	<i>Arabidopsis thaliana</i>		
<i>PAT1</i> and <i>B33</i>	<i>Solanum tuberosum</i> Patatin class I genes	Tuber-specific	β -Carotene expression in potato tubers to improve their nutrition value (Diretto et al. 2007)
ABRE (ABA-responsive element)	Various species	ABA-inducible	Bonetta and McCourt (1998)
<i>alcR</i> and <i>aclA</i> system	<i>Aspergillus nidulans</i> Ethanol-inducible TF and alcohol dehydrogenase genes	Ethanol- and acetaldehyde-inducible	Caddick et al. (1998)
<i>pin2</i>	<i>Solanum tuberosum</i> Proteinase inhibitor II gene	Wound-inducible	Keil et al. (1990)
<i>PtDrl02</i>	<i>Populus tremula</i> Cellulose synthase gene	Xylem-specific	Bidirectional promoters (Zheng et al. 2011)
<i>HaFAD2-1</i>	Sunflower	Seed-specific	Strong expression in seeds equal to <i>CaMV 35S</i> (Zavallo et al. 2010)
<i>PRPI</i> group	Various plants	Pathogen inducible	Pathogen response (Mehrotra et al. 2011)
<i>CaMV 35S</i> core promoter and <i>cis</i> elements	Synthetic	Phytohormones inducible	Fast and easy indication of plant pathogen infections <i>in vivo</i> (Mehrotra et al. 2011)
<i>SK2</i>	<i>Solanum tuberosum</i> Endochitinase gene	Pistil-specific	Ficker et al. (1997)
<i>CHS15</i>	<i>Phaseolus vulgaris</i> Chalcone synthase gene	Petal-specific	Faktor et al. (1996)
<i>TobRB7</i>	<i>Nicotiana tabacum</i> Aquaporin gene	Root-specific	Nematode-resistant tobacco (Yamamoto et al. 1991; Ravichandra 2014)

Table 1 (continued)

Sequence/type	Origin	Function	Application/reference
<i>lat52</i>	<i>Solanum lycopersicum</i> Cysteine-rich protein gene	Pollen-specific	Pollen labelling with orange fluorescent protein in transgenic <i>Nicotiana</i> hybrid plants (Twell et al. 1989; Hollis Rice et al. 2013)
ACC oxidase promoter	<i>Malus sylvestris</i> ACC oxidase gene	Fruit-specific	Atkinson et al. (1998)
Enhancers			
<i>CaMV 35S</i>	<i>Cauliflower mosaic virus</i>	Specific promoters gain enhanced and constitutive activity	Yang et al. (2011)
<i>LAT52</i>	<i>Brassica</i> pollen	Specific promoters gain added enhancer specificity	Construction of enhancer-promoter chimeras (Yang et al. 2011)
<i>AGL5</i>	<i>Brassica</i> carpel		
3'UTR	Maize <i>crtRB1</i> gene	Increasing of the protein translation and accumulation level	β-Carotene production and accumulation in maize seeds (Vignesh et al. 2013)
3'UTR	<i>Oryza sativa</i> <i>SSP</i> genes	Intensification of transcription and translation, mRNA stabilization	Recombinant proteins production and accumulation in seeds (Li et al. 2012)
Insulators			
<i>TBS</i>	<i>Petunia hybrida</i>	Enhancer-promoter insulator	Yang et al. (2011)
<i>EXOB</i>	<i>Aphage</i>		
Signal sequences			
<i>NLS/NES</i>	Synthetic	Directing of a product in/out of nucleus	Sorting of a resistance proteins transformation (Huang et al. 2014)
<i>KDEL</i>	Synthetic	Directing of a product to ER and accumulation	Production of a <i>Dengue virus</i> vaccines and antibodies in <i>N. tabacum</i> and <i>M. citrifolia</i> (Martinez et al. 2011) Production of a phytase in <i>B. napus</i> (Peng et al. 2006) Production of a mouse antibodies 14D9 in <i>N. tabacum</i> (López et al. 2010) Production of a 7Crp vaccine against alergenenes using <i>O. sativa</i> (Takaiwa et al. 2009)

In recent years, there has been a tendency to select and verify putative unknown regulatory sequences in order to use them in synthetic constructs. Comparative studies reported several times higher expression of the mentioned before complete cassettes *Pcec* in comparison to the well-known strong promoter *CaMV 35S* (Koul et al. 2012). Laboratory studies cease to use native sequences in favour of the modified cassettes, which are versatile and cause multiplied effect.

Acknowledgments The authors thank Amadeus Xu (University of Bristol, UK) for editorial suggestions.

Authors' contributions RB and KS drafted the manuscript. KHK and AKK conceived the review and helped to draft the manuscript.

Funding This review was funded by University of Lodz Statutory Funding (Grant number B151100000201.01).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agarwal P, Garg V, Gautam T, Pillai B, Kanoria S, Kumar Burma P (2014) A study on the influence of different promoter and 50UTR (URM) cassettes from *Arabidopsis thaliana* on the expression level of the reporter gene β-glucuronidase in tobacco and cotton. Transgenic Res 23:351–363
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) Molecular biology of the cell, 4th edn. Garland Science, New York
- Allen GC (2008) The role of nuclear matrix attachment regions in plants. In: Nick P (ed) Plant cell monographs. Springer, Berlin, pp 1–29
- An Y-Q, McDowell JM, Huang S, McKinney EC, Chambliss S, Meagher RB (1996) Strong, constitutive expression of the *Arabidopsis* ACT2/ACT8 actin subclass in vegetative tissues. Plant J 10:107–121

- Atkinson RG, Bolitho KM, Wright MA, Iturriagaitia-Bueno T, Reid SJ, Ross GS (1998) Apple ACC-oxidase and polygalacturonase: ripening-specific gene expression and promoter analysis in transgenic tomato. *Plant Mol Biol* 38:449–460
- Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* 69:3613–3634
- Benfey PN, Ren L, Chua N-H (1990) Tissue-specific expression from *CaMV* 35S enhancer subdomains in early stages of plant development. *EMBO J* 9:1677–1684
- Bhullar S, Charkavarthy S, Advani S, Datta S, Pental D, Burma PK (2003) Strategies for development of functionally equivalent promoters with minimum sequence homology for transgene expression in plants: cis-elements in a novel DNA context versus domain swapping. *Plant Physiol* 132:988–998
- Bisht NC, Jagannath A, Gupta V, Kumar Burma P, Pental D (2004) A two gene–two promoter system for enhanced expression of a restorer gene (barstar) and development of improved fertility restorer lines for hybrid seed production in crop plants. *Mol Breed* 14:129–144
- Bisht NC, Jagannath A, Burma PK, Pradhan AK, Pental D (2007) Retransformation of a male sterile barnase line with the barstar gene as an efficient alternative method to identify male sterility restorer combinations for heterosis breeding. *Plant Cell Rep* 26:727–733
- Bock R (2013) Strategies for metabolic pathway engineering with multiple transgenes. *Plant Mol Biol* 83:21–31
- Bonetta D, McCourt P (1998) Genetic analysis of ABA signal transduction pathways. *Trends Plant Sci* 3:231–235
- Borysyuk M, Pogrebnyak N, Andrianov V, Kostenyuk I (2013) Engineered plant biomass for biodiesel and bioethanol production. *Patent WO* 2013/033369A2
- Browning KS, Bailey-Serres J (2015) Mechanism of cytoplasmic mRNA translation. *Arabidopsis Book* 13:e0176
- Buchner P, Rochat C, Willems S, Boutin J-P (2002) Characterization of a tissue-specific and developmentally regulated β -1,3-glucanase gene in pea (*Pisum sativum*). *Plant Mol Biol* 49:171–186
- Bustos MM, Guiltinan MJ, Jordano J, Begum D, Kalkan FA, Hall TC (1989) Regulation of beta-glucuronidase expression in transgenic tobacco plants by an A/T-rich, cis-acting sequence found upstream of a French bean beta-phaseolin gene. *Plant Cell* 1:839–853
- Caddick MX, Greenland AJ, Jepson I, Krause K-P, Qu N, Riddell KV, Salter MG, Schuch W, Sonnwald U, Tomsett AB (1998) An ethanol inducible gene switch for plants used to manipulate carbon metabolism. *Nat Biotechnol* 16:177–180
- Chakraborty S, Chakraborty N, Datta A (2000) Increased nutritive value of transgenic potato by expressing a nonallergenic seed albumin gene from *Amaranthus hypochondriacus*. *Proc Natl Acad Sci USA* 97:3724–3729
- Chamberland S, Daigle N, Bernier F (1992) The legumin boxes and the 3' part of a soybean beta-conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. *Plant Mol Biol* 19:937–949
- Chen X, Wang Z, Wang J, Wang M, Zhao L, Wang G (2007) Isolation and characterization of Brittle2 promoter from *Zea Mays* and its comparison with Ze19 promoter in transgenic tobacco plants. *Plant Cell Tiss Organ Cult* 88:11–20
- Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Mol Biol* 18(4):675–689
- Ciechanover A (1998) The ubiquitin–proteasome pathway: on protein death and cell life. *EMBO J* 17:7151–7160
- Cornejo MJ, Luth D, Blankenship KM, Anderson OD, Blechl AE (1993) Activity of a maize ubiquitin promoter in transgenic rice. *Plant Mol Biol* 23:567–581
- Corrado G, Karali M (2009) Inducible gene expression systems and plant biotechnology. *Biotechnol Adv* 27:733–743
- Daniell H (2006) Production of biopharmaceuticals and vaccines in plants via the chloroplast genome. *Biotechnol J* 1(10):1071–1079
- Dare AP, Schaffer RJ, Lin-Wang K, Allan AC, Hellens RP (2008) Identification of a cis-regulatory element by transient analysis of co-ordinately regulated genes. *Plant Methods* 4:17
- Davière JM, Achard P (2013) Gibberellin signaling in plants. *Development* 140(6):1147–1151
- Deikman J, Fischer RL (1988) Interaction of a DNA binding factor with the 5'-flanking region of an ethylene-responsive fruit ripening gene from tomato. *EMBO J* 7:3315–3320
- Delano-Frier JP, Pearce G, Huffaker A, Stratmann JW (2013) Systemic wound signaling in plants. In long-distance systemic signaling and communication in plants. Springer, Berlin/Heidelberg, pp 323–362
- Deng W, Roberts SG (2005) A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* 19:2418–2423
- Dill A, Thomas SG, Hu J, Steber CM, Sun T (2004) The Arabidopsis F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. *Plant Cell* 16:1392–1405
- Diretto G, Al-Babili S, Tavazza R, Papacchioli V, Beyer P, Giuliano G (2007) Metabolic engineering of potato carotenoid content through tuber-specific overexpression of a bacterial mini-pathway. *PLoS One* 4:e350
- Ebina I, Takemoto-Tsutsumi M, Watanabe S, Koyama H, Endo Y, Kimata K, Igarashi T, Murakami K, Kudo R, Ohsumi A, Noh AL, Takahashi H, Naito S, Onouchi H (2015) Identification of novel *Arabidopsis thaliana* upstream open reading frames that control expression of the main coding sequences in a peptide sequence-dependent manner. *Abbrev Nucleic Acids Res* 43(3):1562–1576
- Faktor O, Kooter JM, Dixon RA, Lamb CJ (1996) Functional dissection of a bean chalcone synthase gene promoter in transgenic tobacco plants. *Plant Mol Biol* 32:849–859
- Ficker M, Wemmer T, Thompson RD (1997) A promoter directing high level expression in pistils of transgenic plants. *Plant Mol Biol* 35:425–431
- Gouiaa S, Khoudi H, Leidi EO, Pardo JM, Masmoudi K (2012) Expression of wheat Na⁺/H⁺ antiporter TNHXS1 and H⁺-pyrophosphatase TVP1 genes in tobacco from a bicistronic transcriptional unit improves salt tolerance. *Plant Mol Biol* 79:137–155
- Guerrero-González ML, Rodríguez-Kessler M, Jiménez-Bremont JF (2014) uORF, a regulatory mechanism of the Arabidopsis polyamine oxidase 2. *Mol Biol Rep* 41:2427–2443
- Hanson J, Hanssen M, Wiese A, Hendriks MMWB, Smeekens S (2008) The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of ASPARAGINE SYNTHETASE1 and PROLINE DEHYDROGENASE2. *Plant J* 53:935–949
- Hauvermale AL, Ariizumi T, Steber CM (2012) Gibberellin signaling: a theme and variations on DELLA repression. *Plant Physiol* 160:83–92
- Hayden CA, Jorgensen RA (2007) Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. *BMC Biol* 5:32
- He ZM, Jiang XL, Qi Y, Luo DQ (2008) Assessment of the utility of the tomato fruit-specific E8 promoter for driving vaccine antigen expression. *Genetica* 133(2):207–214
- Hernandez-Garcia CM, Finer JJ (2014) Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci* 217–218:109–119

- Hily JM, Singer SD, Yang Y, Liu Z (2009) A transformation booster sequence (TBS) from *Petunia hybrida* functions as an enhancer-blocking insulator in *Arabidopsis thaliana*. *Plant Cell Rep* 28:1095–1104
- Hollis Rice J, Millwood RJ, Mundell RE, Chambers OD, Abercrombie LL, Maelor Davies H, Stewart CN (2013) An orange fluorescent protein tagging system for real-time pollen tracking. *BMC Res Notes* 6:383
- Huang YY, Shi Y, Lei Y, Li Y, Fan J, Xu YJ, Ma XF, Zhao JQ, Xiao S, Wang WM (2014) Functional identification of multiple nucleocytoplasmic trafficking signals in the broad-spectrum resistance protein RPW8.2. *Planta* 239:455–468
- Ivanov IP, Atkins JF, Michael AJ (2010) A profusion of upstream open reading frame mechanisms in polyamine-responsive translational regulation. *Nucleic Acids Res* 2(38):353–359
- Jefferson R, Goldsborough A, Bevan M (1990) Transcriptional regulation of a patatin-I gene in potato. *Plant Mol Biol* 14:995–1006
- Jung H, Kim J-K, Ha S-H (2011) Use of animal viral internal ribosome entry site sequence makes multiple truncated transcripts without mediating polycistronic expression in rice. *J Korean Soc Appl Biol Chem* 54(5):678–684
- Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* 339:225–229
- Kadonaga JT (2012) Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol* 1(1):40–51
- Kaufmann K, Pajoro A, Angenent GC (2010) Regulation of transcription in plants: mechanism controlling developmental switches. *Nat Rev Gene* 11:830–842
- Kawaguchi R, Bailey-Serres J (2002) Regulation of translational initiation in plants. *Curr Opin Plant Biol* 5(5):460–465
- Keil M, Sanchez-Serrano J, Schell J, Willmitzer L (1990) Localization of elements important for the wound-inducible expression of a chimeric potato proteinase inhibitor II-CAT gene in transgenic tobacco plants. *Plant Cell* 2:61–70
- Klose H, Günl M, Usadel B, Fischer R, Commandeur U (2013) Ethanol inducible expression of a mesophilic cellulase avoids adverse effects on plant development. *Biotechnol Biofuels* 6(1):1
- Klug WS, Cummings MR (2003) Concepts of genetics. Prentice Hall, Upper Saddle River
- Koltunow AM, Truettner J, Cox KH, Walroth M, Goldberg RB (1990) Different temporal and spatial gene expression patterns occur during anther development. *Plant Cell* 2:1201–1224
- Koul B, Yadav R, Sanyal I, Sawant S, Sharma V, Amla DV (2012) *Cis*-acting motifs in artificially synthesized expression cassette leads to enhanced transgene expression in tomato (*Solanum lycopersicum* L.). *Plant Physiol Biochem* 61:131–141
- Kumar S, AlAbed D, Whitteck JT, Chen W, Bennett S, Asberry A, Wang X, DeSloover D, Rangasamy M, Wright TR, Gupta M (2015) A combinatorial bidirectional and bicistronic approach for coordinated multi-gene expression in corn. *Plant Mol Biol* 87:341–353
- Kutach AK, Kadonaga JT (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol Cell Biol* 20(13):4754–4764
- Kwak H, Lis JT (2013) Control of transcriptional elongation. *Annu Rev Renet* 47:483
- Lewin B (2001) Genes VII. Artmed Editora, Porto Alegre
- Li WJ, Dai LL, Chai ZJ, Yin ZJ, Qu LQ (2012) Evaluation of seed storage protein gene 3'-untranslated regions in enhancing gene expression in transgenic rice seed. *Transgenic Res* 21:545–553
- Lira JM, Wright TR, Hey TD, Strange MTL, Baker LW (2011) Constructs for expressing herbicide tolerance genes, related plants, and related trait combinations. Patent US 2011/0195845A1
- Liu S, Bao Y (2009) Effects of copy number of an octopine synthase enhancer element and its distance from the TATA box on heterologous expression in transgenic tobacco. *Acta Phys Plant* 31:705–710
- Liu X-J, Prat S, Willmitzer L, Frommer WB (1990) *Cis* regulatory elements directing tuber-specific and sucrose-inducible expression of a chimeric class I patatin promoter/GUS-gene fusion. *Mol Gen. Gene* 223:401–406
- Liu W, Mazarei M, Rudis MR, Fethe MH, Stewart CN Jr (2011) Rapid in vivo analysis of synthetic promoters for plant pathogen phyto-sensing. *BMC Biotechnol* 11:108
- Liu W, Yuan JS, Stewart CN Jr (2013) Advanced genetic tools for plant biotechnology. *Nat Rev Genet* 14:781–793
- López J, Lencina F, Petrucci S, Marconi P, Alvarez MA (2010) Influence of the KDEL signal, DMSO and mannitol on the production of the recombinant antibody 14D9 by long-term *Nicotiana tabacum* cell suspension culture. *Plant Cell Tiss Organ Cult* 103:307–314
- Ma S, Wang A (2012) Molecular farming in plants: an overview. In: Ma S, Wang A (eds) *Molecular farming in plants: recent advances and future prospects*. Springer, Berlin/Heidelberg, pp 1–20
- Madanala R, Gupta V, Pandey AK, Srivastava S, Pandey V, Singh PK, Tuli R (2015) Tobacco chloroplasts as bioreactors for the production of recombinant superoxide dismutase in plants, an industrially useful enzyme. *Plant Mol Biol Rep* 33(4):1107–1115
- Mann DGJ, King ZR, Liu W, Joyce BL, Percifield RJ, Hawkins JS, LaFayette PR, Artelt BJ, Burris JN, Mazarei M, Bennetzen JL, Parrott WA, Stewart CN (2011) Switchgrass (*Panicum virgatum* L.) polyubiquitin gene (PvUbi1 and PvUbi2) promoters for use in plant transformation. *BMC Biotechnol* 11:74
- Mariani C, De Beuckeleer M, Truettner J, Leemans J, Goldberg RB (1990) Induction of male sterility in plants by a chimeric ribonuclease gene. *Nature* 347:737–741
- Martinez CA, Giulietti AM, Rodriguez Talou J (2011) Expression of a KDEL-tagged *Dengue virus* protein in cell suspension cultures of *Nicotiana tabacum* and *Morinda citrifolia*. *Plant Cell Tiss Organ Cult* 107:91–100
- McCarter JP (2009) Molecular approaches toward resistance to plant-parasitic nematodes. In: *Cell biology of plant nematode parasitism*. Springer, Berlin/Heidelberg, pp 239–267
- McElroy D, Zhang W, Cao J, Wu R (1990) Isolation of an efficient actin promoter for use in rice transformation. *Plant Cell* 2(2):163–171
- Meagher RB, Li Y (2002) Metal resistant plants and phytoremediation of environmental contamination. Patent WO 2002/048335A2
- Medberry SL, Lockhart BEL, Olszewski NE (1992) The *Commelina yellow mottle virus* promoter is a strong promoter in vascular and reproductive tissues. *Plant Cell* 4:185–192
- Mehrotra R, Gupta G, Sethi R, Bhalothia P, Kumar N, Mehrotra S (2011) Designer promoter: an artwork of *cis* engineering. *Plant Mol Biol* 75:527–536
- Miao Y, Ding Y, Sun Q-Y, Xu Z-F, Jiang L (2008) Plant bioreactors for pharmaceuticals. *Biotechnol Genet Eng Rev* 25:363–380
- Mohammed A, Douches DS, Pett W, Grafius E, Coombs J, Liswidowati J, Li W, Madkour MA (2000) Evaluation of potato tuber moth (*Lepidoptera: Gelechiidae*) resistance in tubers of Bt-cry5 transgenic potato lines. *J Econ Entomol* 93:472–476
- Nyikó T, Sonkoly B, Mérai Z, Benkovics AH, Silhavy D (2009) Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol Biol* 71:367–378
- O'Keefe BR, Murad AM, Vianna GR, Ramessar K, Saucedo CJ, Wilson J, Buckheit KW, da Cunha NB, Araújo ACG, Lacorte CC, Madeira L, McMahon JB, Rech EL (2015) Engineering soya bean seeds as a scalable platform to produce cyanovirin-N, a non-ARV microbicide against HIV. *Plant Biotechnol J* 13:884–892
- Odell JT, Chua N-H (1985) Identification of DNA sequences required for activity of the *Cauliflower mosaic virus* 35S promoter. *Nature* 313:810–812

- Papadakis ED, Nicklin SA, Baker AH, White SJ (2004) Promoters and Control elements: designing expression cassettes for gene therapy. *Curr Gene Ther* 4:89–113
- Park S-H, Yi N, Kim YS, Jeong M-H, Bang S-W, Choi YD, Kim J-K (2010) Analysis of five novel putative constitutive gene promoters in transgenic plants. *J Exp Bot* 61:2459–2467
- Paul W, Hodge R, Smartt S, Draper J, Scott R (1992) The isolation and characterization of the tapetum specific *Arabidopsis thaliana* A9 gene. *Plant Mol Biol* 19:611–622
- Peng R-H, Yao Q-H, Xiong A-S, Cheng Z-M, Li Y (2006) Codon-modifications and an endoplasmic reticulum-targeting sequence additively enhance expression of an *Aspergillus niger* phytase-gene in transgenic canola. *Plant Cell Rep* 25:124–132
- Peremarti A, Twyman RM, Gómez-Galera S, Naqvi S, Farré G, Sabalza M, Miralpeix B, Dashevskaya S, Yuan D, Ramessar K, Christou P, Zhu C, Bassie L, Capell T (2010) Promoter diversity in multigene transformation. *Plant Mol Biol* 73:363–378
- Phillips T (2008) Regulation of transcription and gene expression in eukaryotes. *Nat Educ* 1(1):199
- Porto MS, Pinheiro MPN, Batista VGL, Cavalcanti dos Santos R, de Albuquerque Melo Filho P, de Lima LM (2014) Plant promoters: an approach of structure and function. *Mol Biotechnol* 56:38–49
- Potenza C, Aleman L, Sengupta-Gopalan C (2004) Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. *In Vitro Cell Dev Biol Plant* 40:1–22
- Rabindran S, Stevenson N, Roy G, Fedorkin O, Skarjinskaia M, Ensley B, Yusibov V (2009) Plant produced human growth hormone shows biological activity in rats. *Biotech progress* (25)2:530–534
- Ravichandra NG (2014) Horticultural nematology. Springer, New Delhi
- Roa-Rodríguez C (2003) Promoters used to regulate gene expression. CAMBIA Intellectual Property, Canberra
- Roslan HA, Salter MG, Wood CD, White MRH, Croft KP, Robson F, Coupland G, Doonan J, Laufs P, Tomsett AB, Caddick MX (2001) Characterization of the ethanol-inducible *alc* gene-expression system in *Arabidopsis thaliana*. *Plant J* 28:225–235
- Rushton PJ, Reinstädler A, Lipka V, Lippol B, Somssich IE (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. *Plant Cell* 14:749–762
- Russell PJ (1996) Genetics, 4th edn. Harper Collins College, New York
- Sagor GHM, Berberich T, Tanaka S, Nishiyama M, Kanayama Y, Kojima S, Muramoto K, Kusano T (2016) A novel strategy to produce sweeter tomato fruits with high sugar contents by fruit-specific expression of a single bZIP transcription factor gene. *Plant Biotechnol J* 14:1116–1126
- Salter MG, Paine JA, Riddell KV, Jepson I, Greenland AJ, Caddick MX, Tomsett AB (1998) Characterisation of the ethanol-inducible *alc* gene expression system for transgenic plants. *Plant J* 16:127–132
- Sawant SV, Singh PK, Madanala R, Tuli R (2001) Designing of an artificial expression cassette for high level expression of transgenes in plants. *Theor Appl Genet* 102:635644
- Saxena D, Stewart CN, Altosaar I, Shu Q, Stotzky G (2004) Larvicidal Cry proteins from *Bacillus thuringiensis* are released in root exudates of transgenic *B. thuringiensis* corn, potato, and rice but not of *B. thuringiensis* canola, cotton, and tobacco. *Plant Physiol Biochem* 42(5):383–387
- Shaaltiel J, Bartfeld D, Hashmueli S, Baum G, Brill-Almon E, Galili G, Dym O, Boldin-Adamsky SA, Silman I, Sussman JL, Futerman AH, Aviezer D (2007) Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnol J* 5(5):579–590
- Shamloo-Dashtpagerdi R, Razi H, Aliakbari M, Lindlöf A, Ebrahimi M, Ebrahimi E (2015) A novel pairwise comparison method for in silico discovery of statistically significant *cis*-regulatory elements in eukaryotic promoter regions: application to *Arabidopsis*. *J Theor Biol* 364:364–376
- Sharma AK, Sharma MK (2009) Plants as bioreactors: recent developments and emerging opportunities. *Biotechnol Adv* 27:811–832
- Sharma N, Russell SD, Bhalla PL, Singh MB (2011) Putative *cis*-regulatory elements in genes highly expressed in rice sperm cells. *Res Notes* 4:319
- Shekhawat UKS, Ganapathi TR (2014) Transgenic banana plants over-expressing MusabZIP53 display severe growth retardation with enhanced sucrose and polyphenol oxidase activity. *Plant Cell Tiss Organ Cult* 116:387–402
- Shepherd DN, Dugdale B, Martin DP, Varsani A, Lakay FM, Bezuidenhout ME, Monjane AL, Thomson JA, Dale J, Rybicki EP (2014) Inducible resistance to maize streak virus. *PLoS One* 9(8):e105932
- Singer SD, Liu Z, Cox KD (2012) Minimizing the unpredictability of transgene expression in plants: the role of genetic insulators. *Plant Cell Rep* 31:13–25
- Sirko A, Vaněk T, Góra-Sochacka A, Redkiewicz P (2011) Recombinant cytokines from plants. *Int J Mol Sci* 12(6):3536–3552
- Spangenberg G, Mouradov A, Griffith ME, Martelotto LG (2010) Modification of fructan biosynthesis, increasing plant biomass, and enhancing productivity of biochemical pathways in a plant. Patent AU 2009/291522A1
- Stoger E, Williams S, Keen D, Christou P (1999) Constitutive versus seed specific expression in transgenic wheat: temporal and spatial control. *Transgenic Res* 8:73–82
- Stratmann JW (2003) Long distance run in the wound response—jasmonic acid is pulling ahead. *Trends Plant Sci* 8(6):247–250
- Sunilkumar G, Connell JP, Smith CW, Reddy AS, Rathore KS (2002) Cotton a-globulin promoter: isolation and functional characterization in transgenic cotton, *Arabidopsis*, and tobacco. *Transgenic Res* 11:347–359
- Sunilkumar G, Campbell LM, Puckhaber L, Stipanovic RD, Rathore KS (2006) Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *PNAS* 103(48):18054–18059
- Takaiwa F, Hirose S, Takagi H, Yang L, Wakasa Y (2009) Deposition of a recombinant peptide in ER-derived protein bodies by retention with cysteine-rich prolamins in transgenic rice seed. *Planta* 229:1147–1158
- Tekoah Y, Shulman A, Kizhner T, Ruderfer I (2015) Large-scale production of pharmaceutical proteins in plant cell culture—the protalix experience. *Plant Biotechnol J* 13:1199–1208
- Thalor SK, Berberich T, Lee SS, Yang SH, Zhu XJ, Imai R, Takahashi Y, Kusano T (2012) Deregulation of sucrose-controlled translation of a bZIP-type transcription factor results in sucrose accumulation in leaves. *PLoS One* 7(3):e33111
- Torbert KA, Gopalraj M, Medberry SL, Olszewski NE, Somers DA (1998) Expression of the *Commelina yellow mottle virus* promoter in transgenic oat. *Plant Cell Rep* 17:284–287
- Tran MK, Schultz CJ, Baumann U (2008) Conserved upstream open reading frames in higher plants. *BMC Genom* 9:361
- Twell D, Wing R, Yamaguchi J, McCormick S (1989) Isolation and expression of an anther-specific gene from tomato. *Mol Gen Genet* 217:240–245
- Twyman RM (2003) Growth and development: control of gene expression, regulation of transcription. In: Thomas B, Murphy DJ, Murray B (eds) *Encyclopedia of applied plant sciences*. Elsevier Science, London, pp 558–567
- Twyman RM, Schillberg S, Fischer R (2013) Optimizing the yield of recombinant pharmaceutical proteins in plants. *Curr Pharm Des* 19(31):5486–5494

- Tyurin AA, Kabardaeva KV, Gra OA, Mustafaev OM, Sadowskaya NS, Pavlenko OS, Goldenkova-Pavlov IV (2016) Efficient expression of a heterologous gene in plants depends on the nucleotide composition of mRNA's 5'-region. *Russ J Plant Physiol* 4(63):511–522
- Urwin PE, Moller SG, Lilley CJ, McPherson MJ, Atkinson HJ (1997) Continual green fluorescent protein monitoring of *Cauliflower mosaic virus* 35S promoter activity in nematode-induced feeding cells in *Arabidopsis thaliana*. *Mol Plant Microbe Interact* 10:394–400
- Vaughn JN, Ellingson SR, Mignone F, von Arnim AG (2012) Known and novel post-transcriptional regulatory sequences are conserved across plant families. *RNA* 18:368–384
- Venter M, Botha FC (2010) Synthetic promoter engineering. In: Pua EC, Davey MR (eds) *Plant developmental biology—biotechnological perspectives*. Springer, Berlin/Heidelberg, pp 393–414
- Vignesh M, Nepolean T, Hossain F, Singh K, Gupta HS (2013) Sequence variation in 3'UTR region of crtRB1 gene and its effect on β -carotene accumulation in maize kernel. *J Plant Biochem Biotechnol* 22(4):401–408
- Yamaguchi-Shinozaki K, Shinozaki K (1994) A novel cis-acting element in a *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salinity stress. *Plant Cell* 6:251–264
- Yamamoto YT, Taylor CG, Acedo GN, Cheng C-L, Conkling MA (1991) Characterization of cis-acting sequences regulating root-specific gene expression in tobacco. *Plant Cell* 3:371–382
- Yang Y, Singer SD, Liu Z (2011) Evaluation and comparison of the insulation efficiency of three enhancer-blocking insulators in plants. *Plant Cell Tiss Organ Cult* 105:405–414
- Zavallo D, Lopez Bilbao M, Hopp HE, Heinz R (2010) Isolation and functional characterization of two novel seed-specific promoters from sunflower (*Helianthus annuus* L.) *Plant Cell Rep* 29(3):239–248
- Zhang W, McElroy D, Wu R (1991) Analysis of rice Act1 5'-region activity in transgenic rice plants. *Plant Cell* 3:1155–1165
- Zhang C, Gai Y, Wang W, Zhu Y, Chen X, Jiang X (2008) Construction and analysis of a plant transformation binary vector pBDGG harbouring a bi-directional promoter fusing dual visible reporter genes. *J Genet Genom* 35:245–249
- Zhang D, Nandi S, Bryan P, Pettit S, Nguyen D, Santos M, Huang N (2010) Expression, purification, and characterization of recombinant human transferrin from rice (*Oryza sativa* L.) *Protein Expr Purif* 74(1):69–79
- Zheng X, Deng W, Luo K, Duan H, Chen Y, McAvoy R, Song S, Pei Y, Li Y (2007) The *Cauliflower mosaic virus* (CaMV) 35S promoter sequence alters the level and patterns of activity of adjacent tissue- and organ-specific gene promoters. *Plant Cell Rep* 26:1195–1203
- Zheng H, Lei Y, Lin S, Zhang Q, Zhang Z (2011) Bidirectionalization of a methyl jasmonate-inducible plant promoter. *Biotechnol Lett* 33:387–393
- Zhou F, Roy B, Arnim AG (2010) Translation reinitiation and development are compromised in similar ways by mutations in translation initiation factor eIF3h and the ribosomal protein RPL24. *BMC Plant Biol* 10:193