(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 November 2015 (26.11.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/177659 A1

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/00* (2006.01)

(21) International Application Number:
PCT/IB2015/052503

(22) International Filing Date:
7 April 2015 (07.04.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2014/03606 19 May 2014 (19.05.2014) ZA
14/575,479 18 December 2014 (18.12.2014) US

(71) Applicant: **AZARGEN BIOTECHNOLOGIES (PTY) LTD** [ZA/ZA]; Natural Science Building, Room 2066, Merriman Avenue, Western Cape, 7600 Stellenbosch (ZA).

(72) Inventors: **ZWIEGELAAR, Jacobus Petrus**; 13 Crisanne Street, Jagtershof, Western Cape, 7580 Kuilsrivier (ZA). **VENTER, Mauritz**; Sauvignon Park, No 13 Smuts Avenue, Western Cape, 7130 Somerset West (ZA).

(74) Agent: **EDWARD NATHAN SONNENBERGS INCORPORATED**; 1 North Wharf Square, Loop Street, Foreshore, 8001 Cape Town (ZA).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2015/177659 A1

(54) Title: SYNTHETIC PROMOTER CONSTRUCT FOR TRANSGENE EXPRESSION

(57) Abstract: The invention is directed to synthetic promoter constructs for enhanced transgene expression in plants and to expression cassettes comprising the synthetic promoter constructs. The expression cassettes may include various elements for improved expression, stability of the expressed protein or efficient purification of the expressed protein, including signal sequences, protease cleavage sites for release of the target protein, trafficking peptides for trafficking of the expressed protein to various plant compartments, and/or various tags. The invention further relates to methods of expression of transgenic proteins in plants with the use of the synthetic promoter constructs and expression cassettes.

SYNTHETIC PROMOTERS FOR TRANSGENE EXPRESSION IN PLANTS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of South African Provisional Patent Application No. 2014/03606, filed on May 19, 2014, the entire contents of which are hereby incorporated by reference.

TECHNICAL FIELD

This invention relates to regulation of gene expression in plants.

BACKGROUND

Unreliable and low-level constitutive transgene expression in plants is one of the major challenges in plant biotechnology applications for crop improvement and bio-pharming. Several molecular elements may play a role, but low-level expression of transgenes is primarily due to weak promoters. Currently, the choice of available promoters is limited and the majority of these promoters are based on conventional wild-type promoter elements of which the cauliflower mosaic virus 35S (CaMV 35S) is one of the most widely used promoters.

However, there is a need to drive transgene expression at levels higher than which is capable by the wild-type and even modified versions of CaMV 35S, especially for bio-pharming applications.

SUMMARY OF THE INVENTION

Disclosed herein are synthetic promoter constructs for enhanced transgene expression in plants; expression cassettes comprising the synthetic promoter constructs; and methods of expressing transgenic proteins in plants using the synthetic promoter constructs.

Provided herein is an isolated polynucleotide comprising a regulatory nucleic acid sequence that either (a) has at least 95% or more sequence identity to a polynucleotide sequence set forth in any one of SEQ ID NOs: 1, 3, 4 or 18, or a functionally equivalent fragment thereof, or (b) hybridises under stringent conditions to the complement of a polynucleotide sequence set forth in any one of SEQ ID NOs: 1, 3, 4 or 18.

In some embodiments, the regulatory nucleic acid sequence has 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 4 or 18 or a functionally equivalent fragment thereof. In some embodiments, the the regulatory nucleic acid sequence has 100% sequence identity to any one of SEQ ID NOs: 1, 3, 4 or 18 or a functionally equivalent fragment thereof.

In some embodiments, the regulatory nucleic acid sequence or functionally equivalent fragment thereof contains one or more motifs selected from the group consisting of SORLREP2AT, ASF1MOTIFCAMV, AS1CAMV, ATRICHPSPETE, GT1MOTIFPSRBCS, INRNTPSADB, CTRMCAMV35S, -10PEHVPSBD, LTREATLTI78, CBFHV, DRECRTCOREAT, LTRECOREATCOR15, MYCCONSENSUSAT, LRENPCABE, SORLIP1AT, BOXIIPCCHS, CCA1ATLHCB1, CGF1ATCAB2, GT1CONSENSUS, IBOXCORE, RBCSCONSENSUS, CIACADIANLELHC, GATABOX, SORLIP5AT, and CACTFTPPCA1.

The isolated polynucleotide can further comprise one or more restriction enzyme sites for inserting a second nucleic acid sequence that is heterologous to the regulatory nucleic acid sequence, such that the inserted second nucleic acid sequence would be operably linked to the regulatory nucleic acid sequence, and the regulatory nucleic acid sequence is capable of directing transcription of the second operably linked nucleic acid sequence.

In some embodiments, the second nucleic acid sequence encodes a target protein heterologous to the regulatory nucleic acid sequence and is operably linked to the regulatory nucleic acid sequence. For example, the second nucleotide sequence can encode a target protein useful in the food or beverage industries, the pharmaceutical industry (e.g., as vaccines or therapeutic proteins), in agriculture, or in the chemical industry. In some embodiments, the second nucleotide sequence encodes a therapeutic protein. In some embodiments, the second nucleotide sequence encodes one or more copies of: (A) a pulmonary surfactant protein-B (SP-B) pre-proprotein or a functional fragment or analog thereof; (B) SP-B mature peptide or a functional fragment or analog thereof; or both (A) and (B).

In some embodiments, the isolated polynucleotides further comprise a terminator sequence, for example, a terminator sequence selected from a Nos poly A, a 35S poly A, a RbcS1 terminator or another terminator known to those skilled in the art. The regulatory nucleic acid sequence or functionally equivalent fragment thereof can be operably linked to the terminator sequence.

The isolated polynucleotides can further comprise one or more of the following elements:

(i) a polynucleotide sequence encoding a signal peptide such as the signal peptide region of equistatin or the signal peptide region of *Nicotiana tabacum* thionin (NtSP);

(ii) a polynucleotide sequence encoding a trafficking peptide such as an endoplasmic reticulum (ER)-trafficking peptide e.g., SEKDEL (SEQ ID NO:34), KDEL (SEQ ID NO:35), HDEL (SEQ ID NO:36); an oil body-trafficking peptide (e.g. oleosin), a protein storage vacuole-trafficking peptide (e.g., a Vacuolar Sorting Determinant (VSD) from for example, barley lectin, common bean phaseolin, or soybean β-conglycinin α' subunit); a plastid, including a chloroplast, chromoplast or leucoplast-trafficking peptide (e.g., a peptide capable of interacting with the thylakoid membrane of a plastid such as the chloroplast targeting signal from the small subunit of Rubisco from *Solanum*); or another trafficking peptide known to those skilled in the art;

(iii) a polynucleotide sequence encoding a tag such as polyhistidine, Leptin, late embryogenesis abundant protein (LEA), Lectin, maltose binding protein (MBP) or glutathione S-transferase (GST), or other tags known to those skilled in the art;

(iv) a polynucleotide sequence encoding a marker protein for detection, such as yellow fluorescent protein (YPet), green fluorescent protein (GFP), chloramphenicol acetyl transferase (CAT), luciferase, or others known to those skilled in the art;

(v) a polynucleotide sequence encoding a protease cleavage site, such as any one or more of an enterokinase, chymosin or Tobacco Etch Virus (TEV) protease cleavage site; or

(vi) a psbA regulatory 5′-UTR or 3′ UTR region for targeting a plastid, such as a chloroplast.

Also provided is an expression cassette for expressing proteins in a plant cell or a plant. The expression cassette comprises the isolated polynucleotides described herein. The plant-based expression cassettes may be suitable for *Agrobacterium tumefaciens*-mediated transformation.

Also provided is a plant cell comprising any of the isolated polynucleotides or plant-based expression cassettes described herein. The plant cell may be in a plurality of plant cells in suspension culture, in plant cells in tissue culture, in plant cells in a leaf of a plant or a transgenic plant or any part thereof. The plant cell may be from a monocot or dicot plant. In some embodiments, the plant is a *Nicotiana tobacum* plant.

Also provided are plant progenies or seeds comprising any of the isolated polynucleotides or plant-based expression cassettes described herein.

Also provided herein is a method of expressing a target protein in a plant by introducing the isolated polynucleotides described herein into a plant cell, and expressing the target protein in the plant cell. The method can further include a step of exposing the plant cell to one or more induction stimuli. For example, such induction stimuli may be any one or more of different intensities and periods of light exposure, cold shock, heat shock, induction of drought conditions or hormone induction such as by Abscisic Acid (ABA).

An “isolated” polynucleotide means a polynucleotide that is synthesized or separated from its native environment and present in sufficient quantity to permit

its identification or use. An isolated polynucleotide can be one that is (i) amplified in vitro by, for example, polymerase chain reaction (PCR); (ii) recombinantly produced by cloning; (iii) purified, as by cleavage and gel separation; or (iv) synthesized by, for example, chemical synthesis. An isolated polynucleotide is one which is readily manipulable by recombinant DNA techniques well known in the art. Thus, a nucleotide sequence contained in a vector in which 5' and 3' restriction sites are known or for which polymerase chain reaction (PCR) primer sequences have been disclosed is considered isolated but a nucleotide sequence existing in its native state in its natural host is not. An isolated polynucleotide can be substantially purified, but need not be. In the context of the present invention, polynucleotide, is used interchangeably with the term "nucleic acid."

As used herein, a "regulatory nucleic acid sequence" refers to a sequence of DNA that is usually, but not always, located upstream (5') to a coding sequence, and controls the expression of the coding sequence by providing the recognition for RNA polymerase and/or other factors required for transcription to start at a particular site. An example of a regulatory nucleic acid sequence is a promoter. A promoter comprises a core promoter element, responsible for the initiation of transcription, as well as other regulatory elements that modify gene expression. It is to be understood that nucleotide sequences, located within introns, or 3' of the coding sequence may also contribute to the regulation of expression of a coding sequence.

The term "functionally equivalent fragment" used herein refers to a portion of the regulatory nucleic acid sequence that retains at least 20% (e.g., at least: 30%; 40%; 50%; 60%; 70%; 75%; 80%; 85%; 90%; 95%; 97%; 98%; 99%; or all) of the biological activity of the regulatory nucleic acid sequence in regulating the expression of an operably linked coding sequence. Methods of measuring and comparing the relative regulatory activity of nucleic acid sequences are well known in the art.

As used herein, the term "sequence identity" refers to the degree of identity between any given query sequence and a subject sequence. Percentage of "sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, where the fragment of the nucleotide sequence in the comparison window may comprise additions or deletions (e.g., gaps or overhangs) as compared to the reference sequence (which does not comprise additions or

deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. The output is the percent identity of the subject sequence with respect to the query sequence. Various means for determination of percent sequence identity between two particular polynucleotide sequences are well known to those skilled in the art. For example, one method for use is the Basic Local Alignment Search Tool (BLAST) tool that finds regions of local similarity between sequences (www.ncbi.nlm.nih.gov/blast/).

The term "operably linked" used herein means the transcription or translation of a heterologous nucleotide sequence is under the influence of the regulatory nucleic acid sequence. "Operably linked" is also intended to mean the joining of two nucleotide sequences such that the coding sequence of each DNA fragment remain in the proper reading frame.

A "heterologous" polynucleotide sequence or protein is a sequence or protein that is not naturally operably linked to and regulated by the regulatory nucleic acid sequence.

The term “analog” is used herein to refer to a protein or peptide molecule that structurally and/or functionally resembles a reference protein or peptide molecule, but contains at least one modification, addition, deletion, or substitution of one or more amino acid residues.

“Stringent conditions” or "stringent hybridization conditions" used herein refers to conditions under which a probe will hybridize to its target sequence to a detectably greater degree than to other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences that are 100% complementary to the probe can be identified (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Typically, stringent conditions will be those in which the salt concentration is about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about

30ºC for short probes (e.g., 10 to 50 nucleotides) and at least about 60ºC for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a representative graphical illustration of the promoter expression cassette described herein.

Figure 2 shows a graphical representation of YPet reporter gene expression in 5 leaves analysed from 3 independent transgenic lines.

Figure 3 shows a graphical representation of a comparison between all candidate synthetic promoter cassettes and the two control promoters (C1 and C2), showing transgenic lines of each promoter with the highest reporter gene expression. The dashed line represents the highest C2 (2x35S) fluorescence value measured (4.34 RFU/µg protein). Error bars show SE between three leaves per transgenic

line measured. WT = wild-type, AZprom-NEG = promoterless negative control ('empty vector').

Figure 4 shows the sequence listing for the polynucleotide sequencesdescribed herein.

Figure 5 shows a Western blot analysis for transgenic plant lines for AZ18_1xSP-B with the SeeBlue® Plus2 Pre-Stained protein standard (Life Technologies, LC5925).

Figure 6 shows a Western blot analysis for transgenic plant lines for AZ18_2xSP-B with the SeeBlue® Plus2 Pre-Stained protein standard (Life Technologies, LC5925).

DETAILED DESCRIPTION OF THE INVENTION

Provided herein are synthetic promoter constructs for enhanced transgene expression in plants; expression cassettes comprising the synthetic promoter constructs; and methods of expressing transgenic proteins in plants using the synthetic promoter constructs.

Temporal and/or spatial control of transgene expression in plants is desirable in many agricultural and biopharmaceutical applications. One way of achieving such control is by rational design of the regulatory nucleic acid sequence, for example, the promoter. Promoters can be inducible, constitutive and/or tissue-specific. Promoter engineering applications for targeted control of transgene activity in plants include biotic and abiotic stress tolerance, high-level transgene expression and sensing environmental and/or chemical stimuli. Depending on the biotechnological application,e.g., crop improvement or commercially viable expression of a biopharmaceutical peptide, both high-level constitutive and targeted/inducible transgene expression systems are important.

Strong constitutive transgene expression continues to be a desirable trait. The Cauliflower Mosaic Virus 35S (CaMV 35S) promoter [Benfey and Chua 1990], and recombinant modifications thereof, e.g. native CaMV 35S plus an additional

35S subdomain-B acting as an enhancer, have been used extensively for high level transgene expression in plants [Benfey et al 1990; Bhullar et al. 2003, 2007]. However, currently, the choice and availability of strong constitutive promoters with comparable or even better activity than CaMV 35S is still relatively limited.

The challenges in plant genetic engineering studies,.e.g., promoter availability, insufficient levels of transgene expression, and homology-dependent gene silencing (HDGS), necessitate the engineering and use of novel synthetic promoters [Venter and Botha 2010]. In addition to the design of novel synthetic strong constitutive promoters, specific *cis*-motif architecture within regulatory promoter modules, may allow for predictive applications to confer inducible expression in response to various environmental, physical or chemical stimuli [Venter 2007]. Compared to conventional wild-type promoters from viral and/or plant origin, well-designed synthetic promoters may offer superior transgene expression levels, refine control of transgene (or multiple transgene) expression and limit homology-dependent gene silencing in plant genetic engineering applications.

The applicant has therefore developed 21 candidate synthetic promoters for testing in transgenic plant lines using a stepwise approach by: 1) identifying promoters of genes known to be highly expressed in plants (for example, by literature searches), 2) selecting specific regions in those promoters (regulatory modules) based on the presence of specific plant transcription factor (TF) binding sites (or *cis*-motifs), for example by using a database of known motifs such as PLACE (Higo *et al.*, 1999) or PlantCARE (Lescot *et al.*, 2002), 3) combining different and/or repeated promoter modules to construct synthetic promoters and 4) *in vivo* testing of the candidate synthetic promoter constructs in transgenic plant lines for efficient expression of a transgenic protein.

Examples of regulatory modules selected for synthetic promoter design included:

- Promoters known for high-level transgene expression in plants;
- Genes known to be expressed at high levels in plants;
- Promoters associated with photosynthesis; and
- Plant transcription factor binding sites as detected with PLACE [Higo *et al.*, 1999].

Furthermore, *cis*-motif selections with the use of PLACE [Higo *et al.*, 1999] were primarily based on:

- Light regulation, activation;
- Leaf specific expression;
- Motifs known to act as enhancers;
- Motifs associated with low temperature responsiveness; and
- Motifs from specific promoters, e.g., CaMV 35S and Ribulose-1,5-bisphosphate carboxylase small subunit promoter (RBCss) known to drive high-level transgene expression in plants.

The candidate regulatory modules were designed and/or extended to comprise a majority selection of the above-mentioned *cis*-motifs, which include motifs that function as inducible elements. Accordingly, a person skilled in the art would appreciate that the method may comprise a step of exposing the plant cell to one or more induction stimuli, including different intensities and periods of light exposure, cold shock, heat shock, induction of drought conditions or hormone induction such as by Abscisic Acid (ABA).

The 21 candidate synthetic promoters were then tested in transgenic tobacco plants with the use of a fluorescent reporter protein, YPet (Yellow Fluorescent Protein), although any other reporter protein known to those skilled in the art may be used, such as luciferase, green fluorescent protein (GFP), chloramphenicol acetyl transferase (CAT) and the like. Interestingly, most of the candidate synthetic promoters designed based using an *in silico* rational design strategy did not show marked improvement, or performed worse than the control CaMV 35S promoters. This demonstrates the necessity for performance of *in vivo* testing in order to determine the efficacy of a synthetic promoter construct. Evidence of efficacy *in silico* is not sufficient. However, several first generation transgenic plant lines from various synthetic promoter constructs did show enhanced YPet expression compared to the control CaMV 35S promoters and one of the synthetic promoter constructs in particular showed exceptionally improved protein expression: it increased protein expression about 5 times more than the highest performing control 2xCaMV 35S promoter.

Accordingly, provided herein are isolated polynucleotides comprising a regulatory nucleic acid sequence. The regulatory nucleic acid sequence can be or

contain, a sequence having at least 80% sequence identity (e.g., at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity) to any one of the polynucleotide sequences set forth in SEQ ID NOs: 1, 3, 4, or 18, or a functionally equivalent fragment thereof. In some embodiments, the regulatory nucleic acid sequence has 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 4 or 18 or a functionally equivalent fragment thereof. In some embodiments, the the regulatory nucleic acid sequence has 100% sequence identity to any one of SEQ ID NOs: 1, 3, 4 or 18 or a functionally equivalent fragment thereof.

The regulatory nucleic acid sequence can also be or contain, a sequence that hybridises under stringent conditions to the complement of any one of SEQ ID NOs: 1, 3, 4 or 18. Hybridization conditions are well known to those skilled in the art. “Stringent conditions” or "stringent hybridization conditions" used herein refers to conditions under which a probe will hybridize to its target sequence to a detectably greater degree than to other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences that are 100% complementary to the probe can be identified (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, often less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30ºC for short probes (e.g., 10 to 50 nucleotides) and at least about 60ºC for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary stringent conditions include hybridization in 6X sodium chloride/sodium citrate (SSC) at 37ºC, followed by a wash in 0.5X or 1x SSC at 55 to 60ºC. Exemplary high stringency conditions include hybridization in 6X SSC at 45°C, followed by a wash in 0.2 X SSC, 0.1 % SDS at 65°C. Duration of hybridization is generally less than about 24 hours, usually about 4 to about 12 hours.

In some embodiments, the regulatory nucleic acid sequence or functionally equivalent fragment thereof may contain one or more motifs selected from the group consisting of SORLREP2AT, ASF1MOTIFCAMV, AS1CAMV, ATRICHPSPETE, GT1MOTIFPSRBCS, INRNTPSADB, CTRMCAMV35S, -10PEHVPSBD, LTREATLTI78, CBFHV, DRECRTCOREAT, LTRECOREATCOR15, MYCCONSENSUSAT, LRENPCABE, SORLIP1AT, BOXIIPCCHS, CCA1ATLHCB1, CGF1ATCAB2, GT1CONSENSUS, IBOXCORE, RBCSCONSENSUS, CIACADIANLELHC, GATABOX, SORLIP5AT, and CACTFTPPCA1. It is noted that the motifs provided above are annotated as set out in the PLACE database (Higo *et al.*, 1999), but may have a different annotation if using another database, such as PlantCARE (Lescot *et al.*, 2002). It is also noted that the motifs shown by the sequences depicted in Table 1 can be in the 5'-3' orientation (+) or the reverse 3'-5' orientation (-).

The isolated polynucleotides can further comprise one or more restriction enzyme sites for inserting a second nucleic acid sequence that is heterologous to the regulatory nucleic acid sequence, such that the inserted second nucleic acid sequence would be operably linked to the regulatory nucleic acid sequence, and the regulatory nucleic acid sequence is capable of directing transcription of the second operably linked nucleic acid sequence. In some embodiments, the second nucleic acid sequence encodes a target protein heterologous to the regulatory nucleic acid sequence and is operably linked to the regulatory nucleic acid sequence.

The promoters described herein can be used as strong constitutive promoters for high-level transgene expression in the food and beverage industries, the pharmaceutical industry (e.g., as vaccines or therapeutic proteins), the agriculture industry, or the chemical industry amongst others. It is recognized that any gene of interest can be operably linked to the promoter sequences disclosed herein and expressed in plants.

In one embodiment of the invention, the expressed transgenic protein is a protein useful in the pharmaceutical industry, in particular, a therapeutic protein. A particular embodiment of the invention, the expressed transgenic protein is a pulmonary surfactant protein-B (SP-B) pre-proprotein, or SP-B mature peptide, functional fragment, or analog thereof [Pryhuber 1998].

Accordingly, the second nucleotide sequence can encode a target protein useful in the food or beverage industries, the pharmaceutical industry, in agriculture, or in the chemical industry. In some embodiments, the second nucleotide sequence encodes a therapeutic protein. In some embodiments, the second nucleotide sequence encodes one or more copies of: (A) a pulmonary surfactant protein-B (SP-B) pre-proprotein or a functional fragment or analog thereof; (B) SP-B mature peptide or a functional fragment or analog thereof; or both (A) and (B).

Apart from the target transgenic protein, the regulatory nucleic acid sequence or functionally equivalent fragment thereof can be operably linked to one or more elements known to those skilled in the art, including, but not limited to the following:

(i) a polynucleotide sequence encoding a signal peptide such as the signal peptide region of equistatin or the signal peptide of *Nicotiana tabacum* thionin (NtSP);

(ii) a polynucleotide sequence encoding a trafficking peptide such as an endoplasmic reticulum (ER)-trafficking peptide e.g., SEKDEL (SEQ ID NO:34), KDEL (SEQ ID NO:35), HDEL (SEQ ID NO:36); an oil body-trafficking peptide (e.g. oleosin), a protein storage vacuole-trafficking peptide (e.g., a Vacuolar Sorting Determinant (VSD) from for example, barley lectin, common bean phaseolin, or soybean β-conglycinin α' subunit); a plastid, including a chloroplast, chromoplast or leucoplast-trafficking peptide (e.g., a peptide capable of interacting with the thylakoid membrane of a plastid such as the chloroplast targeting signal from the small subunit of Rubisco from *Solanum*); or another trafficking peptide known to those skilled in the art;

(iii) a polynucleotide sequence encoding a tag such as polyhistidine, Leptin, late embryogenesis abundant protein (LEA), Lectin, maltose binding protein (MBP) or glutathione S-transferase (GST), or other tags known to those skilled in the art for improved protein expression, stability or purification;

(iv) a polynucleotide sequence encoding a marker protein for detection, such as yellow fluorescent protein (YPet), green fluorescent

protein (GFP), chloramphenicol acetyl transferase (CAT), luciferase, or others known to those skilled in the art; and

(v) a polynucleotide sequence encoding a protease cleavage site flanking the target protein, such as any one or more of an enterokinase, chymosin or Tobacco Etch Virus (TEV) protease cleavage site, for release of the target protein of interest by proteolytic cleavage.

Furthermore, the regulatory nucleic acid sequence or functionally equivalent fragment thereof of the invention can be operably linked to a psbA regulatory 5′-UTR and 3′ UTR region for targeting a plastid, in particular a chloroplast.

In some embodiments, the isolated polynucleotides further comprise a terminator sequence, for example, a terminator sequence selected from a Nos poly A, a 35S poly A, a RbcS1 terminator or another terminator known to those skilled in the art. The regulatory nucleic acid sequence or functionally equivalent fragment thereof may be operably linked to the terminator sequence.

Also provided herein are expression cassettes comprising the isolated polynucleotides described herein. The expression cassettes can be suitable for Agrobacterium tumefaciens-mediated transformation. The expression cassette can include a regulatory nucleic acid sequence that either (a) has at least 95% or more sequence identity to a polynucleotide sequence set forth in any one of SEQ ID NOs: 1, 3, 4 or 18, or a functionally equivalent fragment thereof, or (b) hybridises under stringent conditions to the complement of a polynucleotide sequence set forth in any one of SEQ ID NOs: 1, 3, 4 or 18. The expression cassette can further include a second nucleic acid sequence that is operably linked to the regulatory nucleic acid sequence and the regulatory nucleic acid sequence is capable of directing transcription of the second operably linked nucleic acid sequence. The second nucleic acid sequence can encode a target protein useful in the food or beverage industries, the pharmaceutical industry (e.g., as vaccines or therapeutic proteins), in agriculture, or in the chemical industry. In some embodiments, the second nucleotide sequence encodes a therapeutic protein, for example, a pulmonary surfactant protein-B (SP-B) pre-proprotein, or SP-B mature peptide, functional fragment, or analog thereof.

The expression cassette can further comprise one or more of the following elements:

(i) a polynucleotide sequence encoding a signal peptide such as the signal peptide region of equistatin or the signal peptide region of Nicotiana tabacum thionin (NtSP);

(ii) a polynucleotide sequence encoding a trafficking peptide such as an endoplasmic reticulum (ER)-trafficking peptide e.g., SEKDEL (SEQ ID NO:34), KDEL (SEQ ID NO:35), HDEL (SEQ ID NO:36); an oil body-trafficking peptide (e.g. oleosin), a protein storage vacuole-trafficking peptide (e.g., a Vacuolar Sorting Determinant (VSD) from for example, barley lectin, common bean phaseolin, or soybean β-conglycinin α' subunit); a plastid, including a chloroplast, chromoplast or leucoplast-trafficking peptide (e.g., a peptide capable of interacting with the thylakoid membrane of a plastid such as the chloroplast targeting signal from the small subunit of Rubisco from Solanum); or another trafficking peptide known to those skilled in the art;

(iii) a polynucleotide sequence encoding a tag such as polyhistidine, Leptin, late embryogenesis abundant protein (LEA), Lectin, maltose binding protein (MBP) or glutathione S-transferase (GST), or other tags known to those skilled in the art;

(iv) a polynucleotide sequence encoding a marker protein for detection, such as yellow fluorescent protein (YPet), green fluorescent protein (GFP), chloramphenicol acetyl transferase (CAT), luciferase, or others known to those skilled in the art;

(v) a polynucleotide sequence encoding a protease cleavage site, such as any one or more of an enterokinase, chymosin or Tobacco Etch Virus (TEV) protease cleavage site; or

(vi) a psbA regulatory 5′-UTR or 3′ UTR region for targeting a plastid, such as a chloroplast.

Also provided are plant cells comprising any of the isolated polynucleotides or plant-based expression cassettes described herein. The plant cell may be a plurality of plant cells in suspension culture, plant cells in tissue culture, or plant cells in a leaf of a plant, a transgenic plant, or any part thereof. The plant cell may be from a monocot or dicot plant. In some embodiments, the plant is a Nicotiana tobacum plant. Also provided are plant progenies or seeds comprising any of the isolated polynucleotides or plant-based expression cassettes described herein.

Also provided herein are methods of expressing a target protein in a plant by introducing the isolated polynucleotides described herein into a plant cell, and expressing the target protein in the plant cell. The methods can further include a step of exposing the plant cell to one or more induction stimuli. For example, such induction stimuli may be any one or more of different intensities and periods of light exposure, cold shock, heat shock, induction of drought conditions or hormone induction such as by Abscisic Acid (ABA).

The invention will be described by way of the following examples which are not to be construed as limiting in any way the scope of the invention.

EXAMPLES

EXAMPLE 1

1.1. Materials and Methods

1.1.1. Promoter synthesis and cloning in plant transformation vector

Synthesis and sub-cloning of all promoter expression cassettes were conducted by DNA2.0 (www.dna20.com) and Blue Heron Biotech (www.blueheronbio.com). Within ImpactVector 1.1 (Plant Research International, Wageningen; www.pri.wur.nl/UK/products/ImpactVector/) each promoter was synthesized between AscI/NcoI sites, replacing the RbcS1 promoter of *Chrysanthemum morifolium*. Each promoter expression cassette within ImpactVector 1.1 comprised of a promoter (synthetic or unmodified) driving a gene coding for YPet fluorescent protein, synthesized between *Nco*I/*Bgl*II sites, and a

terminator of RbcS1 gene from *C. morifolium* between *Bgl*II/*Pac*I sites (as illustrated in Fig.1). Subsequently, each promoter expression cassette was excised and sub-cloned in a modified (to accommodate *Asc*I and *Pac*I restriction sites) binary vector suitable for *Agrobacterium*-mediated plant transformation.

1.1.2. Synthetic Promoter Design

Regulatory modules for synthetic promoter designs were based on various factors including:

- Promoters known for high-level transgene expression in plants;
- Genes known to be expressed at high levels in plants;
- Promoters associated with photosynthesis; and
- Plant transcription factor binding sites as detected with PLACE [Higo *et al.*, 1999].

In addition, *cis*-motif selections with the use of PLACE [Higo *et al.*, 1999] were primarily based on:

- Light regulation, activation;
- Leaf specific expression;
- Motifs known to act as enhancers;
- Motifs associated with low temperature responsiveness; and
- Motifs from specific promoters i.e. CaMV 35S and Ribulose-1,5-bisphosphate carboxylase small subunit promoter (RBCss) known to drive high-level transgene expression in plants.

The candidate regulatory modules were designed and/or extended to comprise a majority selection of the above-mentioned type *cis*-motifs, which include motifs that function as inducible elements.

Two promoters were chosen as control promoters for comparative analysis to 21 newly designed synthetic promoters. The control promoters were wild-type CaMV 35S promoter (Genbank Acc no. V00140) designated as C1 [SEQ ID NO:22] and a double enhancer CaMV 35S domain-B, also called 2x35S, that has an extension of 73bp at Domain_B (Genbank AC: V00140.1) of the original 35S promoter, together with an additional Domain_B (with 73bp extension) fused to it,

and a *5'-UTR* designated as C2 [SEQ ID NO:23]. The synthetic promoters were designated as AZprom-1 to AZprom-21.

Using the ImpactVector expression cassette and combining a binary vector such as pCAMBIA, specific restriction enzyme sites that should not cut in the promoters for sub-cloning purposes were identified. Where specific restriction enzyme sites were detected, revisions were proposed and sequences submitted to the PLACE database to revise the sequence in such a way that original putative transcription factor-binding sites remained identical to original sequence comprising restriction enzyme-site.

Table 1 shows the various *cis*-motifs present in the promoters selected to drive high-level expression of transgenes in plants. Table 2 shows the various candidate synthetic promoters designed for testing, including their size and components.

1.1.3. Generation of transgenic plants and genomic PCR

Tobacco plants (*Nicotiana tabacum* var Samsun) were maintained on MS medium in a temperature (22 °C) regulated growth room at a 16 hour light/8 hour dark photoperiod. Promoter expression cassettes sub-cloned in a binary vector were mobilized into *Agrobacterium tumefaciens* strain LBA4404 via electroporation and tobacco leaves were transformed using a standard leaf disc method [Horsch et al 1985]. Plantlets were regenerated under antibiotic selection on MS medium and primary transgenic tobacco plantlets were hardened off and grown in a containment glasshouse at 22 °C. After approximately 12 weeks, plants were transferred to a growth room regulated at 24 oC with a 16 hour light/8 hour dark photoperiod at a light intensity of 316 µmol.m-2.s-1. To confirm transgenic status, genomic DNA was extracted from putative transgenic tobacco leaf material using the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific Inc, USA). PCR was performed, using primers: YPet_frw (5'-CTCAGTAAGTGGGGAAGGTGAAGGC-3') [SEQ ID NO: 27] and YPet_rev (5'-TGCCAGCTGAACACCTCCATCCTCG-3') [SEQ ID NO: 28], at a predenaturation cycle of 94 oC for 2 min followed by 35 cycles of amplification (94 °C denaturation, 30 sec; 55 °C annealing, 30 sec; 72 °C polymerization, 30 sec) to amplify a 457 bp

fragment using GoTaq® DNA Polymerase (Promega Corporation, Madison, USA). All PCR reactions were carried out in a Perkin-Elmer GeneAmp® Thermocycler 9700 (Perkin Elmer Corporation, Wellesley, USA) and PCR products were visualized in ethidium bromide-stained 1 % (w/v) agarose gels.

TABLE 1: PLACE *cis*-motif analysis for synthetic promoter design

MOTIF	Associated Function	CORE sequence	Description
-10PEHVPSBD	Circadian/Light	TATTCT	light regulation
AS1CAMV	Leaf	CCACTGACGTAAGGGATGACGCACAATCC (SEQ ID NO:37)	(activation sequence 1) in CaMV 35S
ASF1MOTIFCAMV	Light/Leaf	TGACG	ASF-1 binding site in CaMV 35S
AT1BOX	Light/Leaf	AATATTTTTATT (SEQ ID NO:38)	AT-rich element
ATRICHPSPETE	Enhancer	AATATACTAGTATTATTTACTAAAAAAAATC (SEQ ID NO:39)	AT-rich
BOXIIPCCHS	Light	ACGTGGC	BoxII
CACGTGMOTIF	Light/ Leaf	CACGTG	G-Box
CACTFTPPCA1	Leaf	(T/C)ACT	mesophyll-specific (CACT)
CCA1ATLHCB1	Leaf	AA(C/A)AATCT	Phytochrome
CBFHV	Low Temp/Drought	(A/G)(C/T)CCGAC	CRT/DRE
CIACADIANLELHC	Circadian/Light/Leaf	CAANNNNATC (SEQ ID NO:40)	Region identified for circadian expression
CTRMCAMV35S	Enhancer	TCTCTCTCT	CT-rich motif in CaMV 35S
DRECRTCOREAT	Low Temp/Drought	(A/G)CCGAC	CRT/DRE
GATABOX	Leaf/Light	GATA	GATA motif in CaMV 35S
GBOXLERBCS	Leaf/Light	(A/C)CACGTGGC	rbcS/G-box

GT1MOTIFPSRBCS	Light/Leaf	(G/T)(A/T)GTG(A/G)(A/T)AA(A/T)(A/G)(A/T) (SEQ ID NO:41)	RBCS/BoxII/GT-1
GT1CONSENSUS	Light/Leaf	G(A/G)(A/T)AA(A/T)	GT-1
HY5AT	Light	TGACACGTGGCA (SEQ ID NO:42)	G-box
IBOX	Light/Leaf	GATAAG	I-box/rbcS
IBOXCORE	Light/Leaf	GATAA	I-box
INRNTPSADB	Light	(C/T)TCANT(C/T)(C/T)	'initiator-element'/TATA-less promoter
LBOXLERBCS	Light/Leaf	AAATTAACCAA (SEQ ID NO:43)	rbcS
LREBOXIPCCHS1	Light/Leaf	AACCTAACCT (SEQ ID NO:44)	BoxI
LRENPCABE	Light/Leaf	ACGTGGCA	LRE-element
LTREATLTI78	Low Temp/Leaf	ACCGACA	LTRE
LTRE1HVBLT49	Low Temp/Leaf	CCGAAA	LTRE
LTRECOREATCOR15	Low Temp/Leaf	CCGAC	LTRE
MYCCONSENSUSAT	Drought/Low Temp/Leaf	CANNTG	Leaf/ICE-1
RBCSCONSENSUS	Leaf	AATCCAA	rbcS/G-box/I-box
SORLIP1AT	Light	GCCAC	Phytochrome-reg
SORLIP5AT	Light	GAGTGAG	Phytochrome-reg
SORLREP2AT	Light	ATAAAACGT	Phytochrome-reg
SORLREP3AT	Light	TGTATATAT	Phytochrome-reg
SV40COREENHAN	Enhancer	GTGG(A/T)(A/T)(A/C/T)G	SV40 core enhancer

TABLE 2: Candidate synthetic promoters

Name	Size (bp)*	Description of Components
AZprom_1 [SEQ ID NO:1]	951	Pea plastocyanin promoter region (-784 to -176) (Genbank: X68313)
		Vacuolar Pyrophosphatase VPP (Grapevine) core promoter for PIC assembly (342bp upstream from VPPase ATG-start codon)
AZprom_2 [SEQ ID NO: 2]	1149	3x copies of the wheat cab-1 enhancer (-357 to -89) (Genbank: X05823) plus 3' cytosine residue
		Vacuolar Pyrophosphatase VPP (Grapevine) core promoter for PIC assembly plus native grapevine VPP 5'-UTR region
AZprom_3 [SEQ ID NO: 3]	1039	2x copies of pea plastocyanin enhancer (original 268bp extended to 445bp) (Genbank: X68313)
		CaMV 35S minimal promoter (-90 to + 8)
		pea plastocyanin 5'-UTR (52bp, as calculated from pea plastocyanin TSS at -53bp)
AZprom_4 [SEQ ID NO:4]	1232	2x copies of pea plastocyanin enhancer (original 268bp extended to 445bp) (Genbank: X68313)
		Vacuolar Pyrophosphatase VPP (Grapevine) core promoter for PIC assembly plus native grapevine VPP 5'-UTR region
AZprom_5 [SEQ ID NO:5]	1149	3x copies of the wheat cab-1 enhancer (-357 to -89) (Genbank: X05823) plus 3' cytosine residue in reverse orientation to AZprom_2
		Vacuolar Pyrophosphatase VPP (Grapevine) core promoter for PIC assembly plus native grapevine VPP 5'-UTR region
AZprom_6 [SEQ ID NO:6]	867	4x copies of the spinach plastocyanin enhancer (-259 to -79) (Genbank: X52288)
		Core/minimal-promoter of spinach plastocyanin PLUS native 5'-UTR region
AZprom_7 [SEQ ID NO:7]	1066	4x copies of the spinach plastocyanin enhancer (-259 to -79) (Genbank: X52288)
		Vacuolar Pyrophosphatase VPP (Grapevine) core promoter for PIC assembly plus native grapevine VPP

		5'-UTR region
AZprom_8 [SEQ ID NO:8]	1405	1x copy of CaMV 35S Domain B (-343 to -91) enhancer
		Arabidopsis Plastocyanin promoter region (includes 5'-UTR region) (NC_003070.9_Ath Plastocyanin_NCBI)
AZprom_9 [SEQ ID NO:9]	1424	VPP (M10.A) promoter region plus ASF1MOTIFCAMV, CTRMCAMV35S and VPP promoter region: -1430 to -159 relative to start codon
		CaMV 35S minimal promoter (Domain A: -90 +8)
		Chrysanth rbcS 5'-UTR (AY163904)
AZprom_10 [SEQ ID NO:10]	1004	2x copies of the spinach plastocyanin enhancer (-259 to -79) (Genbank: X52288)
		VPP (M10.A) promoter region plus CBFHV, DRECRTCOREAT, LTREATLTI78, LTRE1HVBLT49, LTRECOREATCOR15, MYCCONSENSUSAT, ASF1MOTIFCAMV, CTRMCAMV35S and VPP promoter region: -642 to start codon
AZprom_11 [SEQ ID NO:11]	895	1x copy of CaMV 35S Domain B (-343 to -91) enhancer
		VPP (M10.A) promoter region plus CBFHV, DRECRTCOREAT, LTREATLTI78, LTRE1HVBLT49, LTRECOREATCOR15, MYCCONSENSUSAT, ASF1MOTIFCAMV, CTRMCAMV35S and VPP promoter region: -642 to start codon
AZprom_12 [SEQ ID NO:12]	1415	VPP (M10.A) promoter region plus ASF1MOTIFCAMV, CTRMCAMV35S and VPP promoter region: -1430 to -159 relative to start codon
		Core/minimal-promoter of spinach plastocyanin PLUS native 5'-UTR region
AZprom_13 [SEQ ID NO:13]	848	2x CaMV 35S Domain B (-343 to -91) enhancer
		Vacuolar Pyrophosphatase VPP (Grapevine) core promoter for PIC assembly plus native grapevine VPP 5'-UTR region
AZprom_14 [SEQ ID NO:14]	1148	2x CaMV 35S Domain B (-343 to -91) enhancer

		VPP (M10.A) promoter region plus CBFHV, DRECRTCOREAT, LTREATLTI78, LTRE1HVBLT49, LTRECOREATCOR15, MYCCONSENSUSAT, ASF1MOTIFCAMV, CTRMCAMV35S and VPP promoter region: -642 to start codon
AZprom_15 [SEQ ID NO:15]	1049	1x CaMV 35S Domain B (-343 to -91) enhancer
		Arabidopsis photosystem I subunit O (PSO-1) promoter region (Genbank: NM_100711)
AZprom_16 [SEQ ID NO:16]	977	1x copy of the spinach plastocyanin enhancer (-259 to -79) (Genbank: X52288)
		Arabidopsis photosystem I subunit O (PSO-1) promoter region (Genbank: NM_100711)
AZprom_17 [SEQ ID NO:17]	743	1x CaMV 35S Domain B (-343 to -91) enhancer
		2x copies of the Ath PSBP-1 selected region (Genbank no. X98108) (-241 to -76 relative to start codon)
		CaMV 35S Domain A minimal promoter (-90 to + 8)
		5'-UTR region of tobacco rbc_ss region (-61bp relative to start codon) (Tobacco_rbcss_EPD: X02353)
AZprom_18 [SEQ ID NO:18]	794	2x copies of the Ath cab2 selected region (-115 relative to TATA-box) (Genbank: X15221)
		Ath cab2 promoter region (-564 relative to start codon)
AZprom_19 [SEQ ID NO:19]	863	2x copies of VPP_mod promoter [SEQ ID NO: 24]
		1x Tobacco core promoter + 5'-UTR (-131bp relative to start codon) (Tobacco_rbcss_EPD: X02353)
AZprom_20 [SEQ ID NO: 20]	934	2x copies of RBCss_mod promoter [SEQ ID NO: 25] (Tobacco_rbcss_EPD: X02353)
		VPP core promoter + 5'-UTR (-131bp relative to start codon)
AZprom_21 [SEQ ID NO: 21]	1400	1x copy of CaMV 35S Domain B (-343 to -91) enhancer
		1x inverted DR region of Ft-PEPc promoter [SEQ ID NO: 26]
		Proximal region (PR) of Ft-PEPc promoter

* without YPet ATG

1.1.4. Fluorescence screening

The YPet reporter is a yellow fluorescent protein (YFP) modified for Forster resonance energy transfer (FRET) applications with an excitation at 517nm and emission at 530nm. Rapid visual screening was conducted using a handheld ROFIN Polilight FLARE PLUS 2 forensic flashlight (ROFIN forensics, Australia) with a Cyan LED output of 485nm to 515nm and orange filter goggles with a range of 190nm to 545nm. Plants transferred to the growth room were continuously screened for 8 to 12 weeks. Highest YPet fluorescence from 35S and 2x35S leaves was used as comparison to screen synthetic promoter plant lines. Plant lines, ranging from 20 cm to 40 cm in height, of which the top 3 leaves had a lower YPet fluorescence were discarded. Plants not discarded were analyzed using a nanodrop fluorospectrometer (Thermo Fisher Scientific Inc, USA).

1.1.5. Sampling

Selected tobacco plants were at least 30cm high (from top of soil in pot). A 10mm diameter cork-bore tool was used to punch 10mm leaf-discs out of tobacco leaves. Each 10mm leaf-disc was punched near the middle of leaf next to, but not including the middle vein, but including a side vein. Five different leaves, designated as L1, L2, L3, L4 and L5 were sampled. The first leaf (L1) was the first leaf of at least 10cm in length. The following leaves going down the plant were then sampled up until the fifth leaf. Each fifth leaf had 5 leaf-discs punched out of the middle of the leaf for analysis.

1.1.6. YPet protein extraction

Extraction buffer used was 1x PBS (Sigma 79383) with 2% 2-mercaptoethanol. Five 10mm leaf-discs (per leaf) per extract were added to 1mL of extraction buffer in 2ml eppendorf tubes. Two 4mm stainless steel balls were placed in each eppendorf tube. The tubes were placed in a tissue-lyzer and shaken at 15min/30rpm. Next, 500µl of chloroform was added. The tubes were then vortexed vigorously for 20 seconds and centrifuged for 10 min at maximum speed (13000 rpm). Supernatent was removed (300mL to 500mL) and used to measure fluorescence intensity.

1.1.7. Fluorescence intensity measurement

The Nanodrop ND3300 protocol for YPet detection (White LED) was used with extraction buffer (1x PBS) as the blank reading. The wild-type plant reading was performed and the blank reading was deducted as background fluorescence. 2 µl of extracts were analysed and the reading was obtained as relative fluorescent unit (RFU) values.

1.1.8. Total protein quantification

Proteins were extracted using the Quant-iT™ Protein Assay kit (Thermo Fisher Scientific Inc, USA) according to manufacturer's instructions. The Nanodrop ND3300 Quant-iT protocol at High-Range was used to produce the standard curve. Blank reading was performed using Quant-iTTM protein buffer.

2µl of extracts were analysed and the reading was obtained as protein quantification (µg/mL). All promoter measurements were quantified as RFU/µg protein.

1.2. Results

Approximately 9 to 28 independent transgenic plant lines were generated for each promoter construct. Fluorometric measurements were made from wild-type tobacco and transgenic lines transformed with an 'empty' expression cassette with YPet reporter gene but no promoter sequence. This was done to analyse background fluorescence and promoter 'leakage' from the binary vector promoter expressing the antibiotic selection gene.

1.2.1. Fluorescence screening with handheld flashlight

First round screening (whole plant analysis) was conducted using the Polilight FLARE flashlight. We started screening all the C1 (35S) and C2 (2x35S) plant lines of which the brightest lines were kept for fluorometric analysis. Plant lines that showed no or very low fluorescence were discarded. Fluoroscence spectroscopy revealed that the threshold brightness for plants to be discarded was

approximately at less than 2 RFU/µg protein per transgenic line. Brightest C1 and C2 control plants were used for visual comparative analysis when transgenic plant lines of all the other promoter expression cassettes were screened with the Polilight.

1.3. Fluorescence spectroscopy

Initial results showed average promoter activity for each leaf position similar to Mitsuhara *et al.* 1996. All promoter constructs had similar expression profiles relative to leaf position (as illustrated in Figure 2A, B and C). However, on average, leaf size for positions L2, L3 and L4 were similar. Based on these profiles we selected the average expression data for leaf positions: L2, L3 and L4 for fluorometric analysis and comparison between all promoter constructs. The primary control promoter for comparative analysis was C2 (2x35S) of which a total of 27 transgenic plant lines were analyzed by fluorescence visulation and spectroscopy. Four promoters were identified as having the highest expression values ranging from 2.71 RFU/µg protein to a maximum of 4.34 RFU/µg protein.

All promoters were screened and transgenic plant lines (T0 primary transformants) with the brightest YPet fluorescence of each candidate synthetic promoter cassette is illustrated in Figure 3. Plant lines with candidate synthetic promoters discarded due to low (< 2 RFU/µg protein) fluorescence are designated as 'n.f.' for 'no fluorescence'. A dashed line on the bar graph (Figure 3) represents the highest C2 (2x35S) value. Transgenic plant lines represented also show the highest reporter gene values for each of the candidate synthetic promoter cassettes. As is evident from the graph AZprom-1, AZprom-3, AZprom-4 and AZprom-18 were the four candidate promoters that worked most efficiently compared with the C2 control. Table 3 indicates the PLACE database motifs identified in each of these four promoter cassettes.

Surprisingly, the AZprom-18 was found to have five times greater expression of the YPet transgene at the highest reporter gene value for a transgenic plant compared with the highest reporter gene value for a transgenic plant with the C2 control promoter. Furthermore, it was also one of the strongest overall synthetic promoters of which every transgenic plant tested had a YPet fluorescence level of above the 2 RFU/µg protein cut-off (Table 4).

Using the handheld Polilight flashlight, transgenic lines could be screened rapidly and lines from several synthetic promoters with overall low activity could be discarded after visual confirmation and comparison to the brightest fluorescence in C1 and C2 plant lines.

The four synthetic promoters AZprom-1, AZprom-3, AZprom-4 and AZprom-18, in particular AZprom-18, identified therefore show utility for use in plant genetic engineering strategies and biopharming applications.

TABLE 3: Motifs identified in selected promoter cassettes by PLACE database analysis

AZprom-1

MOTIF	**POSITION from 5'-start to 3'-end (orientation + or -)**
PEA PLASTOCYANIN PROMOTER (609bp)	
CACTFTPPCA1	19 (+)
CACTFTPPCA1	78 (-)
MYCCONSENSUSAT	108 (+)
MYCCONSENSUSAT	108 (-)
ASF1MOTIFCAMV1	68 (-)
MYCCONSENSUSAT	171 (+)
MYCCONSENSUSAT	171 (-)
MYCCONSENSUSAT	224 (+)
MYCCONSENSUSAT	224 (-)
CACTFTPPCA1	254 (-)
GATABOX	278 (+)
GT1CONSENSUS	436 (+)
GATABOX	441 (-)
CACTFTPPCA1	455 (+)
CACTFTPPCA1	481 (-)
ATRICHPSPETE	496 (+)
CACTFTPPCA1	500 (+)
CACTFTPPCA1	504 (-)
CACTFTPPCA1	513 (+)
CCA1ATLHCB1	520 (+)
SORLIP1AT	530 (+)
BOXIIPCCHS	530 (-)
LTRECOREATCOR15	535 (-)

GATABOX	542 (+)
IBOXCORE	542 (+)
IBOXCORE	581 (+)
CACTFTPPCA1	588 (+)
CACTFTPPCA1	603 (+)

GRAPEVINE VPP PROMOTER (342bp)

RBCSCONSENSUS	634 (-)
SORLIP1AT	658 (-)
GT1CONSENSUS	689 (+)
GT1CONSENSUS	695 (+)
GT1CONSENSUS	699 (-)
SORLREP2AT	720 (-)
GT1MOTIFPSRBCS	741 (-)
GT1CONSENSUS	743 (-)
INRNTPSADB	746 (+)
CACTFTPPCA1	748 (+)
ASF1MOTIFCAMV	759 (-)
GT1CONSENSUS	763 (-)
IBOXCORE	764 (-)
GATABOX	765 (-)
ASF1MOTIFCAMV	768 (-)
CACTFTPPCA1	777 (+)
CTRMCAMV35S	780 (+)
CTRMCAMV35S	782 (+)
CACTFTPPCA1	827 (-)
CACTFTPPCA1	847 (-)
INRNTPSADB	850 (-)
SORLIP5AT	851 (+)
CACTFTPPCA1	852 (-)
-10PEHVPSBD	884 (+)
GATABOX	905 (+)
GT1CONSENSUS	913 (-)
LTREATLTI78	931 (-)
CBFHV	932 (-)
DRECRTCOREAT	932 (-)
LTRECOREATCOR15	932 (-)

AZprom-3

MOTIF	POSITION from 5'-start to 3'-end (orientation + or -)

PEA PLASTOCYANIN PROMOTER (445bp) X2

ASF1MOTIFCAMV	4 (-)
MYCCONSENSUSAT	7 (+)
MYCCONSENSUSAT	7 (-)
MYCCONSENSUSAT	60 (+)
MYCCONSENSUSAT	60 (-)
CACTFTPPCA1	90 (-)
GATABOX	114 (+)
GT1CONSENSUS	272 (+)
GATABOX	277 (-)
CACTFTPPCA1	291 (+)
CACTFTPPCA1	317 (-)
ATRICHPSPETE	332 (+)
CACTFTPPCA1	336 (+)
CACTFTPPCA1	340 (-)
CACTFTPPCA1	349 (+)
CCA1ATLHCB1	356 (+)
SORLIP1AT	366 (+)
BOXIIPCCHS	366 (-)
LTRECOREATCOR15	371 (-)
GATABOX	378 (+)
IBOXCORE	378 (+)
IBOXCORE	417 (+)
CACTFTPPCA1	424 (+)
CACTFTPPCA1	439 (+)
ASF1MOTIFCAMV	449 (-)
MYCCONSENSUSAT	452 (+)
MYCCONSENSUSAT	452 (-)
MYCCONSENSUSAT	505 (+)
MYCCONSENSUSAT	505 (-)
CACTFTPPCA1	535 (-)
GATABOX	559 (+)
GT1CONSENSUS	717 (+)
GATABOX	722 (-)
CACTFTPPCA1	736 (+)
CACTFTPPCA1	762 (-)
ATRICHPSPETE	777 (+)
CACTFTPPCA1	781 (+)
CACTFTPPCA1	785 (-)
CACTFTPPCA1	794 (+)

CCA1ATLHCB1	801 (+)
SORLIP1AT	811 (+)
BOXIIPCCHS	811 (-)
LTRECOREATCOR15	816 (-)
GATABOX	823 (+)
IBOXCORE	823 (+)
IBOXCORE	862 (+)
CACTFTPPCA1	869 (+)
CACTFTPPCA1	884 (+)

CaMV 35S minimal promoter (98bp)

AS1CAMV	894 (+)
CACTFTPPCA1	895 (+)
ASF1MOTIFCAMV	898 (+)
ASF1MOTIFCAMV	910 (+)
CACTFTPPCA1	923 (+)
GATABOX	926 (-)
INRNTPSADB	961 (+)
MYCCONSENSUSAT	968 (+)
MYCCONSENSUSAT	968 (-)

Pea plastocyanin 5'-UTR (52b) as calculated from pea plastocyanin Transcriptional Start Site at -53bp)

CACTFTPPCA1	1007 (-)
CACTFTPPCA1	1025 (+)
GT1CONSENSUS	1034 (+)

AZprom-4

MOTIF **POSITION from 5'-start to 3'-end (orientation + or -)**

PEA PLASTOCYANIN PROMOTER (445bp) X2

ASF1MOTIFCAMV	4 (-)
MYCCONSENSUSAT	7 (+)
MYCCONSENSUSAT	7 (-)
MYCCONSENSUSAT	60 (+)
MYCCONSENSUSAT	60 (-)
CACTFTPPCA1	90 (-)
GATABOX	114 (+)
GT1CONSENSUS	272 (+)

GATABOX	277 (-)
CACTFTPPCA1	291 (+)
CACTFTPPCA1	317 (-)
ATRICHPSPETE	332 (+)
CACTFTPPCA1	336 (+)
CACTFTPPCA1	340 (-)
CACTFTPPCA1	349 (+)
CCA1ATLHCB1	356 (+)
SORLIP1AT	366 (+)
BOXIIPCCHS	366 (-)
LTRECOREATCOR15	371 (-)
GATABOX	378 (+)
IBOXCORE	378 (+)
IBOXCORE	417 (+)
CACTFTPPCA1	424 (+)
CACTFTPPCA1	439 (+)
ASF1MOTIFCAMV	449 (-)
MYCCONSENSUSAT	452 (+)
MYCCONSENSUSAT	452 (-)
MYCCONSENSUSAT	505 (+)
MYCCONSENSUSAT	505 (-)
CACTFTPPCA1	535 (-)
GATABOX	559 (+)
GT1CONSENSUS	717 (+)
GATABOX	722 (-)
CACTFTPPCA1	736 (+)
CACTFTPPCA1	762 (-)
ATRICHPSPETE	777 (+)
CACTFTPPCA1	781 (+)
CACTFTPPCA1	785 (-)
CACTFTPPCA1	794 (+)
CCA1ATLHCB1	801 (+)
SORLIP1AT	811 (+)
BOXIIPCCHS	811 (-)
LTRECOREATCOR15	816 (-)
GATABOX	823 (+)
IBOXCORE	823 (+)
IBOXCORE	862 (+)
CACTFTPPCA1	869 (+)
CACTFTPPCA1	884 (+)

GRAPEVINE VPP PROMOTER (342bp)

RBCSCONSENSUS	915 (-)
SORLIP1AT	939 (-)
GT1CONSENSUS	970 (+)
GT1CONSENSUS	976 (+)
GT1CONSENSUS	980 (-)
SORLREP2AT	1001 (-)
GT1MOTIFPSRBCS	1022 (-)
GT1CONSENSUS	1024 (-)
INRNTPSADB	1027 (+)
CACTFTPPCA1	1029 (+)
ASF1MOTIFCAMV	1040 (-)
GT1CONSENSUS	1044 (-)
IBOXCORE	1045 (-)
GATABOX	1046 (-)
ASF1MOTIFCAMV	1049 (-)
CACTFTPPCA1	1058 (+)
CTRMCAMV35S	1061 (+)
CTRMCAMV35S	1063 (+)
CACTFTPPCA1	1108 (-)
CACTFTPPCA1	1128 (-)
INRNTPSADB	1131 (-)
SORLIP5AT	1132 (+)
CACTFTPPCA1	1133 (-)
-10PEHVPSBD	1165 (+)
GATABOX	1186 (+)
GT1CONSENSUS	1194 (-)
LTREATLTI78	1212 (-)
CBFHV	1213 (-)
DRECRTCOREAT	1213 (-)
LTRECOREATCOR15	1213 (-)

AZprom-18

MOTIF **POSITION from 5'-start to 3'-end (orientation + or -)**

Ath cab2 selected region (115bp) X2

LRENPCABE	18 (-)
SORLIP1AT	19 (+)
BOXIIPCCHS	19 (-)
CACTFTPPCA1	31 (+)

GATABOX	39 (+)
GT1CONSENSUS	39 (+)
IBOXCORE	39 (+)
CCA1ATLHCB1	47 (+)
RBCSCONSENSUS	58 (+)
CGF1ATCAB2	75 (+)
GATABOX	75 (+)
GT1CONSENSUS	75 (+)
IBOXCORE	75 (+)
CACTFTPPCA1	84 (+)
GATABOX	91 (+)
GATABOX	110 (-)
LRENPCABE	133 (-)
SORLIP1AT	134 (+)
BOXIIPCCHS	134 (-)
CACTFTPPCA1	146 (+)
GATABOX	154 (+)
GT1CONSENSUS	154 (+)
IBOXCORE	154 (+)
CCA1ATLHCB1	162 (+)
RBCSCONSENSUS	173 (+)
CGF1ATCAB2	190 (+)
GATABOX	190 (+)
GT1CONSENSUS	190 (+)
IBOXCORE	190 (+)
CACTFTPPCA1	199 (+)
GATABOX	206 (+)
GATABOX	225 (-)

Ath cab2 promoter region (-564 relative to start codon)

ASF1MOTIFCAMV	237 (-)
GATABOX	254 (+)
GATABOX	256 (-)
CIACADIANLELHC	273 (-)
GATABOX	290 (+)
GT1CONSENSUS	290 (+)
IBOXCORE	290 (+)
CACTFTPPCA1	295 (-)
ASF1MOTIFCAMV	297 (+)
CACTFTPPCA1	335 (-)
GT1CONSENSUS	350 (-)

GT1CONSENSUS	359 (+)
GATABOX	368 (+)
GT1CONSENSUS	382 (+)
CACTFTPPCA1	403 (+)
GT1CONSENSUS	438 (-)
CACTFTPPCA1	453 (+)
CACTFTPPCA1	510 (-)
CACTFTPPCA1	514 (-)
CACTFTPPCA1	528 (-)
CACTFTPPCA1	535 (-)
SORLIP1AT	536 (-)
GT1CONSENSUS	576 (+)
MYCCONSENSUSAT	589 (+)
MYCCONSENSUSAT	589 (-)
LRENPCABE	611 (-)
SORLIP1AT	612 (+)
BOXIIPCCHS	612 (-)
CACTFTPPCA1	624 (+)
GATABOX	632 (+)
GT1CONSENSUS	632 (+)
IBOXCORE	632 (+)
CCA1ATLHCB1	640 (+)
RBCSCONSENSUS	651 (+)
CGF1ATCAB2	668 (+)
GATABOX	668 (+)
GT1CONSENSUS	668 (+)
IBOXCORE	668 (+)
CACTFTPPCA1	677 (+)
GATABOX	684 (+)
GATABOX	703 (-)
RBCSCONSENSUS	714 (+)
CIACADIANLELHC	718 (+)
CACTFTPPCA1	721 (+)
GATABOX	724 (-)
CACTFTPPCA1	742 (+)
SORLIP5AT	746 (-)
CACTFTPPCA1	748 (+)

TABLE 4: Number and percentage of transgenic plant lines positive for YPet fluorescence above the 2 RFU/µg protein cut-off

Total No of plant lines	Promoter	Fluorescence spectroscopy +ve	%
19	C1 (35S)	9	47
27	C2 (2x35S)	12	44
12	**AZprom-1**	**4**	**33**
9	AZprom-2	0	/
21	**AZprom-3**	**14**	**67**
27	**AZprom-4**	**7**	**26**
13	AZprom-5	0	/
27	AZprom-6	2	7
28	AZprom-7	0	/
11	AZprom-8	0	/
10	AZprom-9	0	/
24	AZprom-10	1	4
10	AZprom-11	0	/
25	AZprom-12	11	44
10	AZprom-13	0	/
10	AZprom-14	0	/
10	AZprom-15	1	10
25	AZprom-16	3	12
10	AZprom-17	2	20
12	**AZprom-18**	**12**	**100**
20	AZprom-19	1	5
10	AZprom-20	4	40
13	AZprom-21	6	46

EXAMPLE 2

2.1. Materials and Methods

2.1.1. Expression of two fusion proteins using synthetic promoter AZprom_18

The expression cassette within ImpactVector 1.1 comprising the AZprom_18 [SEQ NO.18] synthetic promoter was modified by replacing the single gene coding for YPet fluorescent protein with cDNA coding for: (i) YPet – linker1 - TEV cleavage site - single copy of human mature (8 kDa) pulmonary surfactant protein SP-B peptide [SEQ ID NO: 29] - TEV cleavage site – linker2 - YPet (AZ18_1xSP-B) [SEQ ID NO: 30]; or (ii) YPet – linker1 - TEV cleavage site - double copy of human mature (8 kDa) pulmonary surfactant protein SP-B peptide [SEQ ID NO: 29] - TEV cleavage site – linker2 - YPet (AZ18_2xSP-B) [SEQ ID NO: 31]. The modified AZprom_18 promoter expression cassettes were then excised and sub-cloned in a binary vector, pCAMBIA, suitable for *Agrobacterium*-mediated plant transformation.

2.1.2. Generation of transgenic plants and fluorescence screening

Tobacco plants (*Nicotiana tabacum* var Samsun) were maintained on MS medium in a temperature (22 °C) regulated growth room at a 16 h photoperiod. The pCAMBIA expression vectors comprising fusion cassettes AZ18_1xSP-B and AZ18_2xSP-B were introduced into *Agrobacterium tumefaciens* strain LBA4404 via electroporation [Mattanovich et al 1989] followed by plant transformation using a standard leaf disc method [Horsch et al 1985]. Tobacco plantlets were regenerated under kanamycin selection on MS medium and primary transgenic plantlets were hardened off and grown in a containment glasshouse under standard glasshouse conditions at 22 °C. To evaluate transgenic status of tobacco plants, plantlets were screened using a handheld ROFIN Polilight FLARE PLUS 2 forensic flashlight (ROFIN forensics, Australia) with a Cyan LED output of 485nm to 515nm and orange filter goggles with a range of 190nm to 545nm. Plantlets showing fluorescence were transferred to the growth room and were continuously screened for 8 to 12 weeks.

2.1.3. Protein extraction and Western blot analysis

Proteins were extracted from the plant tissue with a buffer containing 1 x phosphate buffered saline (PBS, Sigma-Aldrich, 79383) with 7M Urea, 2M

Thiourea, 5% CHAPS and 5% 2-mercaptoethanol. 500mg of plant leaf material that was ground up in liquid Nitrogen was extracted with 1ml of extraction buffer. The extract was vortexed for 1 min and left on the bench for 10 minutes where after it was vortexed again and centrifuged for 10 min at 13000 x g. The supernatant was used directly to load into the gel for western blot analysis. SDS PAGE was performed with the Life technologies Bolt™ system (Life Technologies) using Bis-Tris MOPS buffer and 4-12% gradient gel. Transfer and western blot was performed with the Life technologies iBlot® system according to the instructions of the manufacturer. Chromogenic detection was done using the iBlot® chromogenic western detection kit (Life Technologies, IB7410-01). All transgenic plant lines from AZ18_1xSP-B and AZ18_2xSP-B were screened with an antibody specific for GFP mutants (Thermo Fischer Scientific, PA1-28521). To serve as a positive control, fluorescent protein (YPet) was extracted from transgenic plants transformed with promoter expression cassette C2 (2x35S).

2.2. RESULTS & DISCUSSION

2.2.1. Western blot analysis

Western blot analysis showed positive detection of the GFP antibody in several transgenic plant lines for both AZ18_1xSP-B and AZ18_2xSP-B at the expected fusion protein size when compared to the SeeBlue® Plus2 Pre-Stained protein standard (Life Technologies, LC5925) as noted in Figure 5 and Figure 6 respectively. GFP antibody detects positive control, fluorescent protein (YPet), at approximately 28 kDa. Detection of GFP positive control in Figure 5 was faint, but visible. Detection of GFP positive control in Figure 6 was strong, but also showed detection of a larger fragment which could be due to formation of a YPet dimer complex. Nine (9) independent plant lines were analysed for each AZ18-expression cassette. Western blot analysis for AZ18_1xSP-B showed GFP antibody detection at the expected size of approximately 68.4 kDa for 7 transgenic lines (Figure 5) and AZ18_2xSP-B showed GFP antibody detection at the expected size of approximately 77 kDa for 8 transgenic lines (Figure 6).

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

REFERENCES

Benfey PN, Ren L, Chua N-H (1990) Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. EMBO J 9:1685–1696

Bhullar S, Chakravarthy S, Advani S, Datta S, Pental D, Kumar Burma P (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

Bhullar S, Datta S, Advani S, Chakravarthy S, Gautam T, Pental D, Kumar Burma P (2007) Functional analysis of the cauliflower mosaic virus 35S promoter: re-evaluation of the role of subdomains B5, B4, and B2 in promoter activity. Plant Biotechnol J 5:696–708

Higo, K., Y. Ugawa, M. Iwamoto and T. Korenaga (1999) Plant cis-acting regulatory DNA elements (PLACE) database:1999. Nucleic Acids Research Vol.27 No.1 pp. 297-300.

Horsch R, Fry J, Hofmann N, Eichhlotz D, Rogers S, Frayley R (1985) A simple and general method for transferring genes into plants. Science 227(4691): 1229-1231

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rousé P and Rombauts S (2002) PlantCARE, a database of plant *cis*-acting refulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Research 30(1):325-327

Mattanovich D, Ruker F, da Camara Machado A, Laimer M, Regner F, Steinkeliner H, Himmler G, Katinger H (1989) Efficient transformation of Agrobacterium spp. by eletroporation. Nucleic Acids Research 17(16):6747

Mitsuhara I, Ugaki M, Hirochika H, Ohshima M, Murakami T, Gotoh Y, Katayose Y, Nakamura S, Honkura R, Nishimiya S, Ueno K, Mochizuki A, Tanimoto H, Tsugawa H, Otsuki Y, Ohashi Y (1996) Efficient promoter cassettes for enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants. Plant Cell Physiol 37:49–59.

Benfey and Chua N-H (1990) The cauliflower mosaic virus 35S promoter: Combinatorial regulation of transcription in plants. Science 250: 959-966

Pryhuber GS (1998) Regulation and function of pulmonary surfactant protein B. Molecular Genetics and Metabolism 64: 217-228

Venter M (2007) Synthetic promoters: genetic control through cis engineering. Trends Plant Sci. 12, 118–124

Venter M & Botha FC (2010) in Plant Developmental Biology — Biotechnological Perspectives Vol. 2 (eds Pua, E. C. & Davey, M. R.) 393–408 (Springer).

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a regulatory nucleic acid sequence (a) having at least 95% or more sequence identity to a polynucleotide sequence set forth in any one of SEQ ID NOs: 1, 3, 4 or 18, or a functionally equivalent fragment thereof, or (b) which hybridises under stringent conditions to the complement of a polynucleotide sequence set forth in any one of SEQ ID NOs: 1, 3, 4 or 18.

2. The isolated polynucleotide according to claim 1, further comprising one or more restriction enzyme sites for inserting a second nucleic acid sequence heterologous to the regulatory nucleic acid sequence into the polynucleotide, such that the inserted second nucleic acid sequence would be operably linked to the regulatory nucleic acid sequence, wherein the regulatory nucleic acid sequence is capable of directing transcription of the second operably linked nucleic acid sequence.

3. The isolated polynucleotide according to claim 2, comprising the second nucleic acid sequence, wherein the second nucleic acid sequence encodes a target protein heterologous to the regulatory nucleic acid sequence and is operably linked to the regulatory nucleic acid sequence.

4. The isolated polynucleotide according to claim 1, wherein the regulatory nucleic acid sequence has at least 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 4 or 18 or a functionally equivalent fragment thereof.

5. The isolated polynucleotide according to claim 1, wherein the regulatory nucleic acid sequence or functionally equivalent fragment thereof contains any one or more motifs selected from the group consisting of SORLREP2AT, ASF1MOTIFCAMV, AS1CAMV, ATRICHPSPETE, GT1MOTIFPSRBCS, INRNTPSADB, CTRMCAMV35S, -10PEHVPSBD, LTREATLTI78, CBFHV, DRECRTCOREAT, LTRECOREATCOR15, MYCCONSENSUSAT, LRENPCABE, SORLIP1AT, BOXIIPCCHS, CCA1ATLHCB1, CGF1ATCAB2, GT1CONSENSUS,

IBOXCORE, RBCSCONSENSUS, CIACADIANLELHC, GATABOX, SORLIP5AT, and CACTFTPPCA1.

6. The isolated polynucleotide according to claim 3, wherein the second nucleotide sequence encodes a target protein useful in the food or beverage industries, the pharmaceutical industry, in agriculture, or in the chemical industry.

7. The isolated polynucleotide according to claim 3, wherein the second nucleotide sequence encodes a therapeutic protein.

8. The isolated polynucleotide according to claim 3, wherein the second nucleotide sequence encodes one or more copies of: (A) a pulmonary surfactant protein-B (SP-B) pre-proprotein or a functional fragment or analog thereof; (B) SP-B mature peptide or a functional fragment or analog thereof; or (C) both (A) and (B).

9. The isolated polynucleotide according to claim 1, further comprising a polynucleotide sequence encoding a signal peptide selected from the group consisting of the signal peptide region of equistatin and of *Nicotiana tabacum* thionin (NtSP).

10. The isolated polynucleotide according to claim 1, further comprising a polynucleotide sequence encoding a trafficking peptide selected from the group consisting of an endoplasmic reticulum (ER)-trafficking peptide; an oil body-trafficking peptide, a protein storage vacuole-trafficking peptide; and a plastid-trafficking peptide.

11. The isolated polynucleotide according to claim 1, further comprising a polynucleotide sequence encoding a tag selected from the group consisting of polyhistidine, Leptin, late embryogenesis abundant protein (LEA), Lectin, maltose binding protein (MBP) and glutathione S-transferase (GST).

12. The isolated polynucleotide according to claim 1, further comprising a polynucleotide sequence encoding a marker protein for detection, selected from

the group consisting of yellow fluorescent protein (YPet), green fluorescent protein (GFP), chloramphenicol acetyl transferase (CAT), and luciferase.

13. The isolated polynucleotide according to claim 1, further comprising a polynucleotide sequence encoding a protease cleavage site, selected from the group consisting of an enterokinase, a chymosin and a Tobacco Etch Virus (TEV) protease cleavage site.

14. The isolated polynucleotide according to claim 1, further comprising a psbA regulatory 5′-UTR and 3′ UTR region for targeting a plastid.

15. An expression cassette for expression of a protein in a plant comprising the isolated polynucleotide according to claim 1.

16. The expression cassette according to claim 15, comprising one or more restriction enzyme sites for inserting a second nucleic acid sequence heterologous to the regulatory nucleic acid sequence into the polynucleotide, such that the inserted second nucleic acid sequence would be operably linked to the regulatory nucleic acid sequence, wherein the regulatory nucleic acid sequence is capable of directing transcription of the second operably linked nucleic acid sequence.

17. The expression cassette according to claim 16, wherein the expression cassette comprises the second nucleic acid sequence, and wherein the second nucleic acid sequence encodes a target protein heterologous to the regulatory nucleic acid sequence and is operably linked to the regulatory nucleic acid sequence.

18. The expression cassette according to claim 17, wherein the second nucleotide sequence encodes one or more copies of: (A) a pulmonary surfactant protein-B (SP-B) pre-proprotein or a functional fragment or analog thereof; (B) SP-B mature peptide or a functional fragment or analog thereof; or (C) both (A) and (B).

19. The expression cassette according to claim 15, further comprising any one or more elements from the group consisting of:

(i) a polynucleotide sequence encoding a signal peptide from the group consisting of the signal peptide region of equistatin and of *Nicotiana tabacum* thionin (NtSP);

(ii) a polynucleotide sequence encoding a trafficking peptide from the group consisting of an endoplasmic reticulum (ER)-trafficking peptide; an oil body-trafficking peptide, a protein storage vacuole-trafficking peptide; and a plastid-trafficking peptide;

(iii) a polynucleotide sequence encoding a tag from the group consisting of polyhistidine, Leptin, late embryogenesis abundant protein (LEA), Lectin, maltose binding protein (MBP) and glutathione S-transferase (GST);

(iv) a polynucleotide sequence encoding a marker protein for detection from the group consisting of yellow fluorescent protein (YPet), green fluorescent protein (GFP), chloramphenicol acetyl transferase (CAT), and luciferase;

(v) a polynucleotide sequence encoding a protease cleavage site from the group consisting of an enterokinase, a chymosin and a Tobacco Etch Virus (TEV) protease cleavage site; and

(vi) a psbA regulatory 5′-UTR and 3′ UTR region for targeting a plastid.

20. A plant cell comprising the isolated polynucleotide according to claim 3.

21. A method of expressing a target protein in a plant comprising introducing the isolated polynucleotide according to claim 3 into a plant cell, and expressing the target protein in the plant cell.

22. The method according to claim 21, further comprising a step of exposing the plant cell to one or more induction stimuli.

23. The method according to claim 22, wherein the one or more induction stimuli are selected from the group consisting of different intensities and periods of light exposure, cold shock, heat shock, induction of drought conditions and hormone induction.

24. The method of claim 23, wherein the hormone induction comprises induction with Abscisic Acid (ABA).

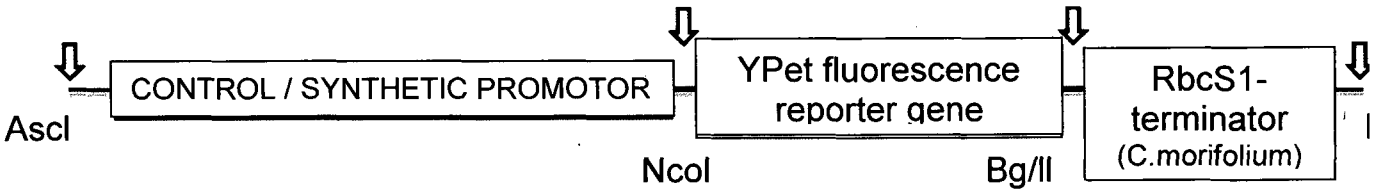

CONTROL / SYNTHETIC PROMOTOR
YPet fluorescence reporter gene
RbcS1-terminator (C.morifolium)
AscI
NcoI
Bg/II

FIG. 1

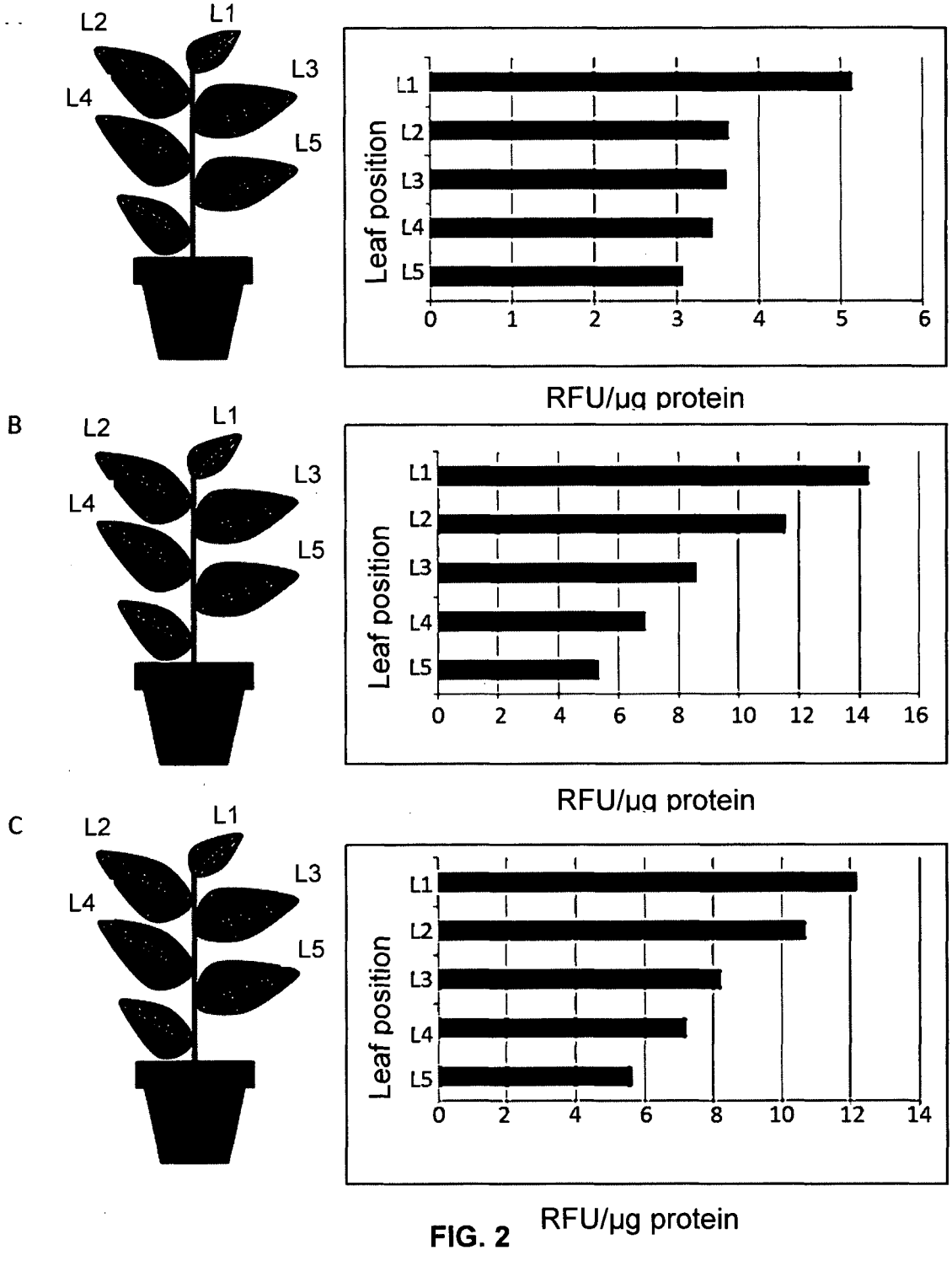

L1
L2
L3
L4
L5
Leaf position
0 1 2 3 4 5 6
RFU/µg protein
B
0 2 4 6 8 10 12 14 16
RFU/µg protein
C
0 2 4 6 8 10 12 14
RFU/µg protein

FIG. 2

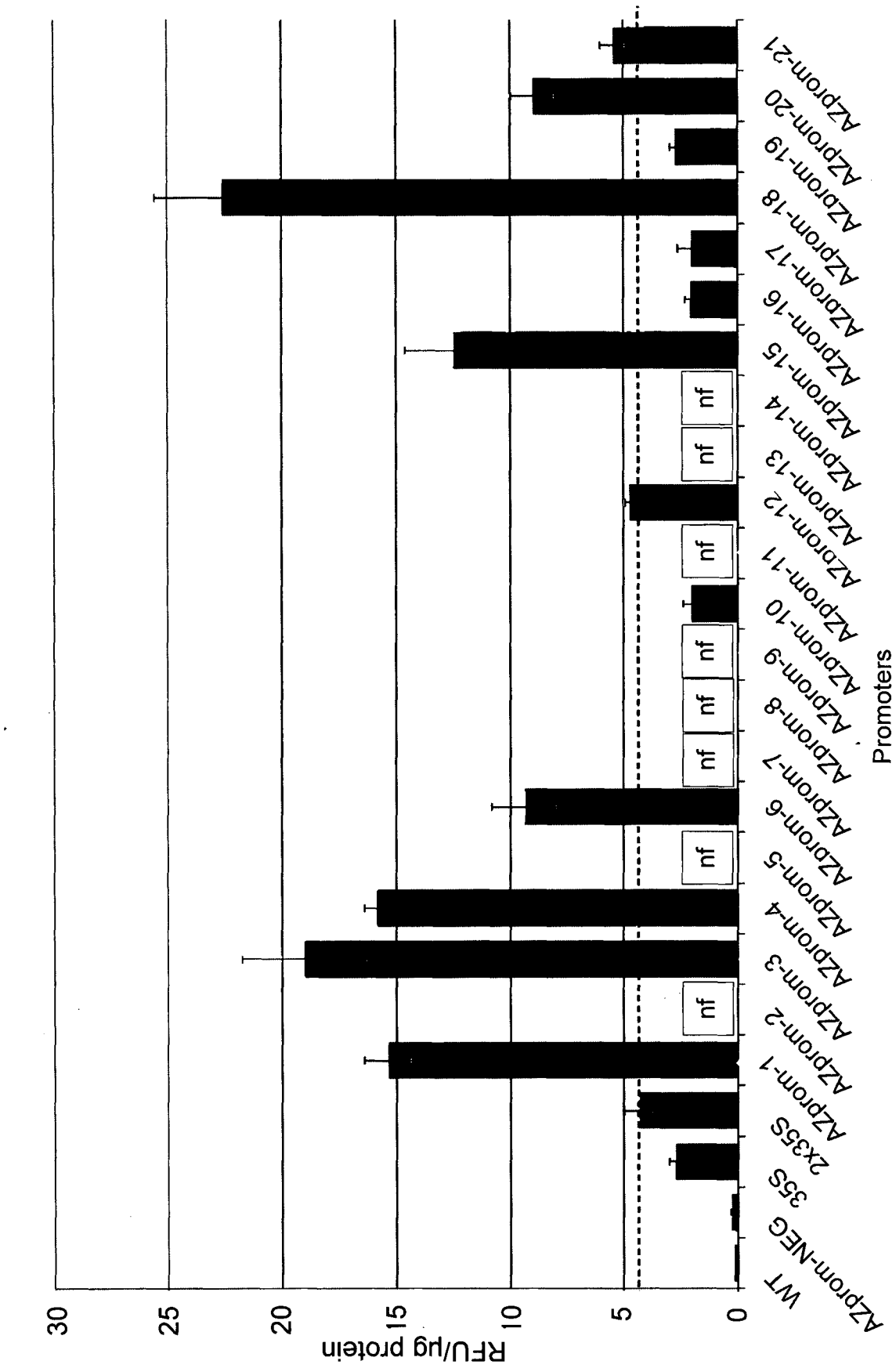

RFU/µg protein
0
5
10
15
20
25
30
WT
AZprom-NEG
35S
2x35S
AZprom-1
AZprom-2
AZprom-3
AZprom-4
AZprom-5
AZprom-6
AZprom-7
AZprom-8
AZprom-9
AZprom-10
AZprom-11
AZprom-12
AZprom-13
AZprom-14
AZprom-15
AZprom-16
AZprom-17
AZprom-18
AZprom-19
AZprom-20
AZprom-21
nf
Promoters

FIG. 3

AZprom_1 [SEQ ID NO:1]
TATGCAACTTACAACGTGCACTCGCGGAGGATTGGACGTGTGCAACTTACAAC
GTACGCATTGTTCGTTCATACAATAGTGTAGAATTGGAGATGTGCAACTTACAA
CATGTGCAACTTACAACGTGCGCTCGCGGAGGAATGTGAAGTTGAACACGTAC
AAACTTACGTCATTTGTGCATGCAGAAGCATAGAGCTGAGCACACAATTCATAA
TTTGAAGGACACATGATTTGCTATAAAGAACTCTTTAGAAGTACCACAACTTTG
ACTGAGTTTGATATAGCTAATAAAGATGGAGCTCATTATAATTTGAATGGCATA
ATGCAAGCTAAACGAACAAGCTTAGTTAATCATGTTAAACAACAATTCTTTGTAA
TAATAAAATTGTCTTTCAACTAGTCCAAGTTTATGAGTTGATTCTTCGGAATAAA
TTAGAAAATATCTTAGATTTTATACTTCATTGATTATTTCATAGAGCAAGTAGGA
GAAATAAAAATATACTAGTATTATTTACTAAAAAAAATCTAAGCCACGTCGGAG
GATAACATCCAACCCAGCCAATCACAGCAATGTTCATCAGATAACCCACTTTAA
GCCCACGCACTCTGAGATCCAAAAATGGTTGAGCTGTTTTGGATTTGGCAATT
AATTGCATCGTGGCAACGTGGAATTAACAAAAATGGAGCTGGAAATGGTAATT
TTCAAAAATATTTTGTAAACGTTTTATAATAATAGAATTATTTTTTCACTCTCTCA
TCGTCATTATCGTCATCATCACTCTCTCTCTCTACGCTTGCATATATATAAACCA
TTGCAGAGCCGCGGAGTGTCAAGCATCGTGGTGGAGTAGAGTGAGAGAACCG
AAGCCAAAGGCAACTCCATTATTCTCTCTCGTGTTCTCGTGATATTGGTTTTCC
GGCGCCGGAGCTTGTCGGTCCGTCCGTCCGTCC

AZprom_2 [SEQ ID NO:2]
ATCGAATTTGTTGGGCAAGCGCCCAGTGTGATGTACGTTGGGAAAACTTGCAA
GAGGATGCGACCAAATGAACTGGTAAACCATCCCGTGAGCGTGGCCTACACA
TTTTAAGCCAGCGGTCTCTTTCGACTTGTCTTACAAAAGCTGGTCCAGTCACGA
GCCTTAGCCCTAACCATAGCCACAAGTACATCTCATCCATTTAAGGCCTCTGC
GTGCACCAATGGCATCCAAGCTGCAGATTTCTTTTCACCACCGTCTCTCTTGTC
AGCATCGAATTTGTTGGGCAAGCGCCCAGTGTGATGTACGTTGGGAAAACTTG
CAAGAGGATGCGACCAAATGAACTGGTAAACCATCCCGTGAGCGTGGCCTAC
ACATTTTAAGCCAGCGGTCTCTTTCGACTTGTCTTACAAAAGCTGGTCCAGTCA
CGAGCCTTAGCCCTAACCATAGCCACAAGTACATCTCATCCATTTAAGGCCTC
TGCGTGCACCAATGGCATCCAAGCTGCAGATTTCTTTTCACCACCGTCTCTCTT
GTCAGCATCGAATTTGTTGGGCAAGCGCCCAGTGTGATGTACGTTGGGAAAAC
TTGCAAGAGGATGCGACCAAATGAACTGGTAAACCATCCCGTGAGCGTGGCC
TACACATTTTAAGCCAGCGGTCTCTTTCGACTTGTCTTACAAAAGCTGGTCCAG
TCACGAGCCTTAGCCCTAACCATAGCCACAAGTACATCTCATCCATTTAAGGC
CTCTGCGTGCACCAATGGCATCCAAGCTGCAGATTTCTTTTCACCACCGTCTC
TCTTGTCAGCAGATCCAAAAATGGTTGAGCTGTTTTGGATTTGGCAATTAATTG
CATCGTGGCAACGTGGAATTAACAAAAATGGAGCTGGAAATGGTAATTTTCAA
AAATATTTTGTAAACGTTTTATAATAATAGAATTATTTTTTCACTCTCTCATCGTC
ATTATCGTCATCATCACTCTCTCTCTCTACGCTTGCATATATATAAACCATTGCA
GAGCCGCGGAGTGTCAAGCATCGTGGTGGAGTAGAGTGAGAGAACCGAAGC
CAAAGGCAACTCCATTATTCTCTCTCGTGTTCTCGTGATATTGGTTTTCCGGCG
CCGGAGCTTGTCGGTCCGTCCGTCCGTCC

AZprom_3 [SEQ ID NO:3]
TTACGTCATTTGTGCATGCAGAAGCATAGAGCTGAGCACACAATTCATAATTTG
AAGGACACATGATTTGCTATAAAGAACTCTTTAGAAGTACCACAACTTTGACTG
AGTTTGATATAGCTAATAAAGATGGAGCTCATTATAATTTGAATGGCATAATGC
AAGCTAAACGAACAAGCTTAGTTAATCATGTTAAACAACAATTCTTTGTAATAAT
AAAATTGTCTTTCAACTAGTCCAAGTTTATGAGTTGATTCTTCGGAATAAATTAG
AAAATATCTTAGATTTTATACTTCATTGATTATTTCATAGAGCAAGTAGGAGAAA
TAAAAATATACTAGTATTATTTACTAAAAAAAATCTAAGCCACGTCGGAGGATAA
CATCCAACCCAGCCAATCACAGCAATGTTCATCAGATAACCCACTTTAAGCCC
ACGCACTCTGTTACGTCATTTGTGCATGCAGAAGCATAGAGCTGAGCACACAA
TTCATAATTTGAAGGACACATGATTTGCTATAAAGAACTCTTTAGAAGTACCAC
AACTTTGACTGAGTTTGATATAGCTAATAAAGATGGAGCTCATTATAATTTGAAT
GGCATAATGCAAGCTAAACGAACAAGCTTAGTTAATCATGTTAAACAACAATTC
TTTGTAATAATAAAATTGTCTTTCAACTAGTCCAAGTTTATGAGTTGATTCTTCG
GAATAAATTAGAAAATATCTTAGATTTTATACTTCATTGATTATTTCATAGAGCAA
GTAGGAGAAATAAAAATATACTAGTATTATTTACTAAAAAAAATCTAAGCCACGT
CGGAGGATAACATCCAACCCAGCCAATCACAGCAATGTTCATCAGATAACCCA
CTTTAAGCCCACGCACTCTGTCTCCACTGACGTAAGGGATGACGCACAATCCC
ACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGA
GGAcacgctgTTCAAACACATACAAATTCAGTAGAGAAGAAACTCATTACTCTTGA
GAAAAA

AZprom_4 [SEQ ID NO:4]
TTACGTCATTTGTGCATGCAGAAGCATAGAGCTGAGCACACAATTCATAATTTG
AAGGACACATGATTTGCTATAAAGAACTCTTTAGAAGTACCACAACTTTGACTG
AGTTTGATATAGCTAATAAAGATGGAGCTCATTATAATTTGAATGGCATAATGC
AAGCTAAACGAACAAGCTTAGTTAATCATGTTAAACAACAATTCTTTGTAATAAT
AAAATTGTCTTTCAACTAGTCCAAGTTTATGAGTTGATTCTTCGGAATAAATTAG
AAAATATCTTAGATTTTATACTTCATTGATTATTTCATAGAGCAAGTAGGAGAAA
TAAAAATATACTAGTATTATTTACTAAAAAAAATCTAAGCCACGTCGGAGGATAA
CATCCAACCCAGCCAATCACAGCAATGTTCATCAGATAACCCACTTTAAGCCC
ACGCACTCTGTTACGTCATTTGTGCATGCAGAAGCATAGAGCTGAGCACACAA
TTCATAATTTGAAGGACACATGATTTGCTATAAAGAACTCTTTAGAAGTACCAC
AACTTTGACTGAGTTTGATATAGCTAATAAAGATGGAGCTCATTATAATTTGAAT
GGCATAATGCAAGCTAAACGAACAAGCTTAGTTAATCATGTTAAACAACAATTC
TTTGTAATAATAAAATTGTCTTTCAACTAGTCCAAGTTTATGAGTTGATTCTTCG
GAATAAATTAGAAAATATCTTAGATTTTATACTTCATTGATTATTTCATAGAGCAA
GTAGGAGAAATAAAAATATACTAGTATTATTTACTAAAAAAAATCTAAGCCACGT
CGGAGGATAACATCCAACCCAGCCAATCACAGCAATGTTCATCAGATAACCCA
CTTTAAGCCCACGCACTCTGAGATCCAAAAATGGTTGAGCTGTTTTGGATTTG
GCAATTAATTGCATCGTGGCAACGTGGAATTAACAAAAATGGAGCTGGAAATG
GTAATTTTCAAAAATATTTTGTAAACGTTTTATAATAATAGAATTATTTTTTCACT
CTCTCATCGTCATTATCGTCATCATCACTCTCTCTCTCTACGCTTGCATATATAT
AAACCATTGCAGAGCCGCGGAGTGTCAAGCATCGTGGTGGAGTAGAGTGAGA
GAACCGAAGCCAAAGGCAACTCCATTATTCTCTCTCGTGTTCTCGTGATATTG
GTTTTCCGGCGCCGGAGCTTGTCGGTCCGTCCGTCCGTCC

AZprom_5 [SEQ ID NO:5]
GCTGACAAGAGAGACGGTGGTGAAAAGAAATCTGCAGCTTGGATGCCATTGG
TGCACGCAGAGGCCTTAAATGGATGAGATGTACTTGTGGCTATGGTTAGGGCT
AAGGCTCGTGACTGGACCAGCTTTTGTAAGACAAGTCGAAAGAGACCGCTGG
CTTAAAATGTGTAGGCCACGCTCACGGGATGGTTTACCAGTTCATTTGGTCGC
ATCCTCTTGCAAGTTTTCCCAACGTACATCACACTGGGCGCTTGCCCAACAAA
TTCGATGCTGACAAGAGAGACGGTGGTGAAAAGAAATCTGCAGCTTGGATGC
CATTGGTGCACGCAGAGGCCTTAAATGGATGAGATGTACTTGTGGCTATGGTT
AGGGCTAAGGCTCGTGACTGGACCAGCTTTTGTAAGACAAGTCGAAAGAGAC
CGCTGGCTTAAAATGTGTAGGCCACGCTCACGGGATGGTTTACCAGTTCATTT
GGTCGCATCCTCTTGCAAGTTTTCCCAACGTACATCACACTGGGCGCTTGCCC
AACAAATTCGATGCTGACAAGAGAGACGGTGGTGAAAAGAAATCTGCAGCTTG
GATGCCATTGGTGCACGCAGAGGCCTTAAATGGATGAGATGTACTTGTGGCTA
TGGTTAGGGCTAAGGCTCGTGACTGGACCAGCTTTTGTAAGACAAGTCGAAAG
AGACCGCTGGCTTAAAATGTGTAGGCCACGCTCACGGGATGGTTTACCAGTTC
ATTTGGTCGCATCCTCTTGCAAGTTTTCCCAACGTACATCACACTGGGCGCTT
GCCCAACAAATTCGATAGATCCAAAAATGGTTGAGCTGTTTTGGATTTGGCAAT
TAATTGCATCGTGGCAACGTGGAATTAACAAAAATGGAGCTGGAAATGGTAAT
TTTCAAAAATATTTTGTAAACGTTTTATAATAATAGAATTATTTTTTCACTCTCTC
ATCGTCATTATCGTCATCATCACTCTCTCTCTCTACGCTTGCATATATATAAACC
ATTGCAGAGCCGCGGAGTGTCAAGCATCGTGGTGGAGTAGAGTGAGAGAACC

GAAGCCAAAGGCAACTCCATTATTCTCTCTCGTGTTCTCGTGATATTGGTTTTC
CGGCGCCGGAGCTTGTCGGTCCGTCCGTCCGTCC

AZprom_6 [SEQ ID NO:6]
GATCGGGTCAGTTTTTGACGGCTCTACTTCAACCCCTCCCATTTCAAAGTTCCA
CATCATCCGATATAATCCGAGCACAACCAATGAAAAACTAGATAATACCCTTTA
TTGGCCCACACATCACAAAAATCCTTTAATCCAATGCCACTAAAAATCCCACAA
ATGAAAACCACACAAAACCGATCGGGTCAGTTTTTGACGGCTCTACTTCAACC
CCTCCCATTTCAAAGTTCCACATCATCCGATATAATCCGAGCACAACCAATGAA
AAACTAGATAATACCCTTTATTGGCCCACACATCACAAAAATCCTTTAATCCAAT
GCCACTAAAAATCCCACAAATGAAAACCACACAAAACCGATCGGGTCAGTTTTT
GACGGCTCTACTTCAACCCCTCCCATTTCAAAGTTCCACATCATCCGATATAAT
CCGAGCACAACCAATGAAAAACTAGATAATACCCTTTATTGGCCCACACATCAC
AAAAATCCTTTAATCCAATGCCACTAAAAATCCCACAAATGAAAACCACACAAA
ACCGATCGGGTCAGTTTTTGACGGCTCTACTTCAACCCCTCCCATTTCAAAGTT
CCACATCATCCGATATAATCCGAGCACAACCAATGAAAAACTAGATAATACCCT
TTATTGGCCCACACATCACAAAAATCCTTTAATCCAATGCCACTAAAAATCCCA
CAAATGAAAACCACACAAAACCATGTAAACAGACATCACCTCATCTATCCTCCA
TTTTATCACCTCTTATTAAATCCCATCCTCGCTCTCAGCACTCTTTGGCAATTGT
CATTTCTTAATTGCATTCCACTTAAACCCCTCCAAAATCCTCCTCCTTCCAATAC
A

AZprom_7 [SEQ ID NO:7]
GATCGGGTCAGTTTTTGACGGCTCTACTTCAACCCCTCCCATTTCAAAGTTCCA
CATCATCCGATATAATCCGAGCACAACCAATGAAAAACTAGATAATACCCTTTA
TTGGCCCACACATCACAAAAATCCTTTAATCCAATGCC
ACTAAAAATCCCACAAATGAAAACCACACAAAACCGATCGGGTCAGTTTTTGAC
GGCTCTACTTCAACCCCTCCCATTTCAAAGTTCCACATCATCCGATATAATCCG
AGCACAACCAATGAAAAACTAGATAATACCCTTTATTGGCCCACACATCACAAA
AATCCTTTAATCCAATGCCACTAAAAATCCCACAAATGAAAACCACACAAAACC
GATCGGGTCAGTTTTTGACGGCTCTACTTCAACCCCTCCCATTTCAAAGTTCCA
CATCATCCGATATAATCCGAGCACAACCAATGAAAAACTAGATAATACCCTTTA
TTGGCCCACACATCACAAAAATCCTTTAATCCAATGCCACTAAAAATCCCACAA
ATGAAAACCACACAAAACCGATCGGGTCAGTTTTTGACGGCTCTACTTCAACC
CCTCCCATTTCAAAGTTCCACATCATCCGATATAATCCGAGCACAACCAATGAA
AAACTAGATAATACCCTTTATTGGCCCACACATCACAAAAATCCTTTAATCCAAT
GCCACTAAAAATCCCACAAATGAAAACCACACAAAACCAGATCCAAAAATGGTT
GAGCTGTTTTGGATTTGGCAATTAATTGCATCGTGGCAACGTGGAATTAACAAA
AATGGAGCTGGAAATGGTAATTTTCAAAAATATTTTGTAAACGTTTTATAATAAT
AGAATTATTTTTTCACTCTCTCATCGTCATTATCGTCATCATCACTCTCTCTCTC
TACGCTTGCATATATATAAACCATTGCAGAGCCGCGGAGTGTCAAGCATCGTG
GTGGAGTAGAGTGAGAGAACCGAAGCCAAAGGCAACTCCATTATTCTCTCTCG
TGTTCTCGTGATATTGGTTTTCCGGCGCCGGAGCTTGTCGGTCCGTCCGTCCG
TCC

AZprom_8 [SEQ ID NO:8]
TGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGA
CAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATTTAAAAGTAAAA
TTCAGACCTCAATAATTCCATCACTTAAACTCATAAGTAGGCACAGGAGGAGCA
TCAAAGCTCTCCAAAATGTGAGTTCTGCTCTCTGAATGATGATCCTCTGTTTCT
TTCTTCTTCTTAAACAAACTCCCAAACCAAGATACCGAACCAGAAGAAGAAGAA
GAAGTCTCCGGTACCGGTACGGACTGACCAGATATCTGACCCATACCATATGG
TAAAGAAGTAGCTGCATGATCCTCCATGAACTGCTCGTTATGCGCAGTACTCT
GAACTTTGTTTAACATGATACCAACTCCTTCTATCATAGCCAACAACACACCTC
CGACTAAAGCCGATCTAGCAGACGCGCCTAAGCCTTGACGCAACGAGAGGAA
TCCTCCAGTGGCTGCACCAGATAAGATCGAGTTCCATCCATCTTCCTTTTGTCT
TGCGTACACCAACGCACAATCGAAGGTTGAGTAAAGACCACCCCACACGGAG
AAGCTTCCTCCGCTTCTCGGCCCGCTCATTCTCAAAGCTTGAACGCCGCCGGA
GAGACGAGCTCCGCCGGGAGAGTTGTAGATTCCTCTTATGAGGTGATACGCT
GATCCACCAACAGCACCCATCGCAAACGCACCTCCGACATCATCTAGGATCCG
ATCCGGACATGGCTCTCTCGATGATTCTGGAGTTCCCATTGATTTTCGATTTTG
ATCTACAAGTAACTTTTTAGGGTATAAGAAAAGAGCCCTTGAAAACGAAACTAA
GAATGAGGAAAAATTACAAAAATGGAAAAATATCGTGTTGTTCACGTTACTTGA
CCCTGGTCAAAGCCCATTAAGGCCCATTTGTAAACACGCTTTATCACAAAATAC
TAATATAAAAACGCCACGTCATTCAAGAACCTTATCTTTATTTAAAACGTGGACA
AACATCCTCCACCTCATAATGCCATGTCACGACACAGTATCTAAAATCAACCAA
TCACAACGCGTCTTTATAGATAACTTGTTTTTTTATGGAGTTTGCTTTTAGAGCC
ATCCATTGTCCTATCTCACTTTCTCTCTTTCACCACATAAAAACTCATAAACTCG
ATCGAACCAAAGCTAAACGAAAAACTTAAAACCCAAATCTTATCACTACTCTAA
AAGAATCTTCATC

AZprom_9 [SEQ ID NO:9]
CAATCCGTCATATTTTTTCAACCTTAAAAATATTTATTATGGGTAAAAATACCTTA
ATATTTTTTTTATAAAAACAATTTTTTATTTAAATACATATGTAAATAATAAAAATA
TTAAAAAATATTTATATAAAAAATGAGTTATTCTTCAAATTCAAAAATAAAAAATA
TATCTAAAATGCAAGTTTGACCTAATAAATGCAACGTCATCATTTTTTAAAATAT
TTTTTACAGAATCAAATTAAAAGAATTCTTATTTTTTCTCATTTTAGAACAAGATA
ACAATAATAATCTTATTTTAAAAAATACTTCATATATTAAAAATTACTATATATATA
CCCCATTTTGATATTTTTTATCATTTATATTGACAAAAAATATTTAAAAAATATTTT
ATTTATTCGTATTTTTAATAACATTTTTAATCAAATAAAAGTAGACAGCGTAGCA
TAACATGAAATCTTGAAAACATTGCGTGAGAAGGAAGAATTAGGATAAGGAACT
GGGAAAGAGTCCAAATGTCAAAAGCACATCTAAAAACATTAAAGCTAAAGATTC
TGTAACTTTTTTATCATTATTTTTGTTCGATTCTATCTTCTACCATTTTGTTGATA
ACATTTACTTAAAAGAATCATAAAGCAAGGTGCCTTAAAAATCAATCACATAATC
GAAACATTCAGGCATTTTGTTGTTTATGAACTTGAATAAATGTTCCCATTAATCA
TGGATCAAATGTGTCACCATGCAAATGTGTGATCTTGAAATTCTGCCACGAAAA
GAGGATAAAGTGATAAGGAATATGGCCAGATCAGCTTCATTTTTAAGGTAATG
GCGATCTTTAGACAGGACGTGGCATCTCTTTTAATTTTTGGTCGGA

TATGTATGGTAAGTGATAATATATATATTTTGATATTATTCTTTGAGAATGTTGTT
ACCAACTTACCGAAGATCACCCACCGTGTCATGTGTGTAGCTCTGCCTCCCAT
AATTATGTAATCAAAAGAGGCCACGTCAACAAATGATTTGACTAACAACCGAAG
ATTTAAAAGAAAGGAATAAATTCGTAAATTAAGGGGTGTTTGTGCAAATAGATC
CAAAAATGGTTGAGCTGTTTTGGATTTGGCAATTAATTGCATCGTGGCAACGTG
GAATTAACAAAAATGGAGCTG

GAAATGGTAATTTTCAAAAATATTTTGTAAACGTTTTATAATAATAGAATTATTTT
TTCACTCTCTCATCGTCATTATCGTCATCATCACTCTCTCTCTCTACGTCTCCAC
TGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTA
TATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTGAGGAACTAATTCACTCA
TTGGATTCATAGAAGTCCATTCCTCCTAAGTATCTAAA

AZprom_10 [SEQ ID NO:10]
GATCGGGTCAGTTTTTGACGGCTCTACTTCAACCCCTCCCATTTCAAAGTTCCA
CATCATCCGATATAATCCGAGCACAACCAATGAAAAACTAGATAATACCCTTTA
TTGGCCCACACATCACAAAAATCCTTTAATCCAATGCCACTAAAAATCCCACAA
ATGAAAACCACACAAAACCGATCGGGTCAGTTTTTGACGGCTCTACTTCAACC
CCTCCCATTTCAAAGTTCCACATCATCCGATATAATCCGAGCACAACCAATGAA
AAACTAGATAATACCCTTTATTGGCCCACACATCACAAAAATCCTTTAATCCAAT
GCCACTAAAAATCCCACAAATGAAAACCACACAAAACCGATAAGGAATATGGC
CAGATCAGCTTCATTTTTAAGGTAATGGCGATCTTTAGACAGGACGTGGCATCT
CTTTTAATTTTTGGTCGGATATGTATGGTAAGTGATAATATATATATTTTGATATT
ATTCTTTGAGAATGTTGTTACCAACTTACCGAAGATCACCCACCGTGTCATGTG
TGTAGCTCTGCCTCCCATAATTATGTAATCAAAAGAGGCCACGTCAACAAATGA
TTTGACTAACAACCGAAGATTTAAAAGAAAGGAATAAATTCGTAAATTAAGGGG
TGTTTGTGCAAATAGATCCAAAAATGGTTGAGCTGTTTTGGATTTGGCAATTAA
TTGCATCGTGGCAACGTGGAATTAACAAAAATGGAGCTGGAAATGGTAATTTT
CAAAAATATTTTGTAAACGTTTTATAATAATAGAATTATTTTTTCACTCTCTCATC
GTCATTATCGTCATCATCACTCTCTCTCTCTACGCTTGCATATATATAAACCATT
GCAGAGCCGCGGAGTGTCAAGCATCGTGGTGGAGTAGAGTGAGAGAACCGA
AGCCAAAGGCAACTCCATTATTCTCTCTCGTGTTCTCGTGATATTGGTTTTCCG
GCGCCGGAGCTTGTCGGTCCGTCCGTCCGTCC

AZprom_11 [SEQ ID NO:11]
TGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGA
CAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATGATAAGGAATA
TGGCCAGATCAGCTTCATTTTTAAGGTAATGGCGATCTTTAGACAGGACGTGG
CATCTCTTTTAATTTTTGGTCGGATATGTATGGTAAGTGATAATATATATATTTT
GATATTATTCTTTGAGAATGTTGTTACCAACTTACCGAAGATCACCCACCGTGT
CATGTGTGTAGCTCTGCCTCCCATAATTATGTAATCAAAAGAGGCCACGTCAA
CAAATGATTTGACTAACAACCGAAGATTTAAAAGAAAGGAATAAATTCGTAAAT
TAAGGGGTGTTTGTGCAAATAGATCCAAAAATGGTTGAGCTGTTTTGGATTTGG
CAATTAATTGCATCGTGGCAACGTGGAATTAACAAAAATGGAGCTGGAAATGG
TAATTTTCAAAAATATTTTGTAAACGTTTTATAATAATAGAATTATTTTTTCACTCT
CTCATCGTCATTATCGTCATCATCACTCTCTCTCTCTACGCTTGCATATATATAA
ACCATTGCAGAGCCGCGGAGTGTCAAGCATCGTGGTGGAGTAGAGTGAGAGA
ACCGAAGCCAAAGGCAACTCCATTATTCTCTCTCGTGTTCTCGTGATATTGGTT
TTCCGGCGCCGGAGCTTGTCGGTCCGTCCGTCCGTCC

AZprom_12 [SEQ ID NO:12]
CAATCCGTCATATTTTTTCAACCTTAAAAATATTTATTATGGGTAAAAATACCTTA
ATATTTTTTTTTATAAAAACAATTTTTTATTTAAATACATATGTAAATAATAAAAATA
TTAAAAAATATTTATATAAAAAATGAGTTATTCTTCAAATTCAAAAATAAAAAATA
TATCTAAAATGCAAGTTTGACCTAATAAATGCAACGTCATCATTTTTTAAAATAT
TTTTTACAGAATCAAATTAAAAGAATTCTTATTTTTTCTCATTTTAGAACAAGATA
ACAATAATAATCTTATTTTAAAAAATACTTCATATATTAAAAATTACTATATATATA
CCCCATTTTGATATTTTTTATCATTTATATTGACAAAAAATATTTAAAAAATATTTT
ATTTATTCGTATTTTTAATAACATTTTTAATCAAATAAAAGTAGACAGCGTAGCA
TAACATGAAATCTTGAAAACATTGCGTGAGAAGGAAGAATTAGGATAAGGAACT
GGGAAAGAGTCCAAATGTCAAAAGCACATCTAAAAACATTAAAGCTAAAGATTC
TGTAACTTTTTTATCATTATTTTTGTTCGATTCTATCTTCTACCATTTTGTTGATA
ACATTTACTTAAAAGAATCATAAAGCAAGGTGCCTTAAAAATCAATCACATAATC
GAAACATTCAGGCATTTTGTTGTTTATGAACTTGAATAAATGTTCCCATTAATCA
TGGATCAAATGTGTCACCATGCAAATGTGTGATCTTGAAATTCTGCCACGAAAA
GAGGATAAAGTGATAAGGAATATGGCCAGATCAGCTTCATTTTTAAGGTAATG
GCGATCTTTAGACAGGACGTGGCATCTCTTTTAATTTTTGGTCGGATATGTATG
GTAAGTGATAATATATATATTTTGATATTATTCTTTGAGAATGTTGTTACCAACTT
ACCGAAGAT
CACCCACCGTGTCATGTGTGTAGCTCTGCCTCCCATAATTATGTAATCAAAAGA
GGCCACGTCAACAAATGATTTGACTAACAACCGAAGATTTAAAAGAAAGGAATA
AATTCGTAAATTAAGGGGTGTTTGTGCAAATAGATCCAAAAATGGTTGAGCTGT
TTTGGATTTGGCAATTAATTGCATCGTGGCAACGTGGAATTAACAAAAATGGAG
CTG
GAAATGGTAATTTTCAAAAATATTTTGTAAACGTTTTATAATAATAGAATTATTTT
TTCACTCTCTCATCGTCATTATCGTCATCATCACTCTCTCTCTCTACGATGTAAA
CAGACATCACCTCATCTATCCTCCATTTTATCACC

TCTTATTAAATCCCATCCTCGCTCTCAGCACTCTTTGGCAATTGTCATTTCTTAA
TTGCATTCCACTTAAACCCCTCCAAAATCCTCCTCCTTCCAATACA

AZprom_13 [SEQ ID NO:13]
TGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGA
CAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATTGAGACTTTTCA
ACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTC
ACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCAT
TGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAA
AGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACC
ACGTCTTCAAAGCAAGTGGATTGATGTGATAGATCCAAAAATGGTTGAGCTGTT
TTGGATTTGGCAATTAATTGCATCGTGGCAACGTGGAATTAACAAAAATGGAG
CTGGAAATGGTAATTTTCAAAAATATTTTGTAAACGTTTTATAATAATAGAATTAT
TTTTTCACTCTCTCATCGTCATTATCGTCATCATCACTCTCTCTCTCTACGCTTG
CATATATATAAACCATTGCAGAGCCGCGGAGTGTCAAGCATCGTGGTGGAGTA
GAGTGAGAGAACCGAAGCCAAAGGCAACTCCATTATTCTCTCTCGTGTTCTCG
TGATATTGGTTTTCCGGCGCCGGAGCTTGTCGGTCCGTCCGTCCGTCC

AZprom_14 [SEQ ID NO:14]
TGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGA
CAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATTGAGACTTTTCA
ACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTC
ACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCAT
TGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAA
AGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACC
ACGTCTTCAAAGCAAGTGGATTGATGTGATGATAAGGAATATGGCCAGATCAG
CTTCATTTTTAAGGTAATGGCGATCTTTAGACAGGACGTGGCATCTCTTTTAAT
TTTTGGTCGGATATGTATGGTAAGTGATAATATATATATTTTGATATTATTCTTT
GAGAATGTTGTTACCAACTTACCGAAGATCACCCACCGTGTCATGTGTGTAGC
TCTGCCTCCCATAATTATGTAATCAAAAGAGGCCACGTCAACAAATGATTTGAC
TAACAACCGAAGATTTAAAAGAAAGGAATAAATTCGTAAATTAAGGGGTGTTTG
TGCAAATAGATCCAAAAATGGTTGAGCTGTTTTGGATTTGGCAATTAATTGCAT
CGTGGCAACGTGGAATTAACAAAAATGGAGCTGGAAATGGTAATTTTCAAAAAT
ATTTTGTAAACGTTTTATAATAATAGAATTATTTTTTCACTCTCTCATCGTCATTA
TCGTCATCATCACTCTCTCTCTCTACGCTTGCATATATATAAACCATTGCAGAG
CCGCGGAGTGTCAAGCATCGTGGTGGAGTAGAGTGAGAGAACCGAAGCCAAA
GGCAACTCCATTATTCTCTCTCGTGTTCTCGTGATATTGGTTTTCCGGCGCCG
GAGCTTGTCGGTCCGTCCGTCCGTCC

AZprom_15 [SEQ ID NO:15]
TGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGA
CAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATGATTTCCATTGA
CGGAGGGCAAAATCGTTGGTGAGACCGAGCTCCAAAAGACCGGTTCCGTCAT
CAAGTACCAGACAATTATCGCCGCCTCCACCGGAGAAGGGGACCAAAACGCC
CTGCAACGCAAGACTTGAAGTCATCGAAACAGGAATTAACAATTTAGCTATAGA
CAATGAGAGGATGATAGATTGGGAACCTGTAACCAAGCACGTTGAAAAAGTAC
ACCGCCTAGGGAGGCGGCGTTCTGAGAAGGCGTCGGTTTAGCGTCCCGGAG
CTGAGAACAAAGCATCTTCACCGCCGCTAAGCTGTAGTCCATTTTACCCGCCG
ATCCGACTCGGCTGATCAATTTCACAAATGCCTTTACCTCATCGATTACCGGAT
TAAACCGAACCGAACAAATGGTATCGGTACAGTAATTTGATTTTGATTTTTGGC
AGATATTGGGAGAGGGAATATTGTGTTTGAAAAAAAAAAAATGGAATTCATGGT
GTTTTGGCTTTTTGACTAATTAAAATCATTAACATGCAAACTAAAAGATGTAAAA
ATCTTTATAAAATACAAACAAAAAAAATAAACAAATTTTGAGACAAAAAATAAAT
CTGGAATCAATCTTATCCAAGAAATAGACCCACAATATCAAAGAAAATAATGAG
GTGTCACAAAGCTATTGATTGATAACACATCCTCGAAATCTCTCTCTTTGTCCA
CAGAATAAAAATCTATTTCTTGTGTTTGTGTTGTATTTACTTAAAAAAATAAAGG
AGATCAGTTTGGGAGAAAGCAGCAAAGAAGAAAA

AZprom_16[SEQ ID NO:16]
GATCGGGTCAGTTTTTGACGGCTCTACTTCAACCCCTCCCATTTCAAAGTTCCA
CATCATCCGATATAATCCGAGCACAACCAATGAAAAACTAGATAATACCCTTTA
TTGGCCCACACATCACAAAAATCCTTTAATCCAATGCCACTAAAAATCCCACAA
ATGAAAACCACACAAAACCGATTTCCATTGACGGAGGGCAAAATCGTTGGTGA
GACCGAGCTCCAAAAGACCGGTTCCGTCATCAAGTACCAGACAATTATCGCCG
CCTCCACCGGAGAAGGGGACCAAAACGCCCTGCAACGCAAGACTTGAAGTCA
TCGAAACAGGAATTAACAATTTAGCTATAGACAATGAGAGGATGATAGATTGG
GAACCTGTAACCAAGCACGTTGAAAAAGTACACCGCCTAGGGAGGCGGCGTT
CTGAGAAGGCGTCGGTTTAGCGTCCCGGAGCTGAGAACAAAGCATCTTCACC
GCCGCTAAGCTGTAGTCCATTTTACCCGCCGATCCGACTCGGCTGATCAATTT
CACAAATGCCTTTACCTCATCGATTACCGGATTAAACCGAACCGAACAAATGGT
ATCGGTACAGTAATTTGATTTTGATTTTTGGCAGATATTGGGAGAGGGAATATT
GTGTTTGAAAAAAAAAAAATGGAATTCATGGTGTTTTGGCTTTTTGACTAATTAA
AATCATTAACATGCAAACTAAAAGATGTAAAAATCTTTATAAAATACAAACAAAA
AAAATAAACAAATTTTGAGACAAAAAATAAATCTGGAATCAATCTTATCCAAGAA
ATAGACCCACAATATCAAAGAAAATAATGAGGTGTCACAAAGCTATTGATTGAT
AACACATCCTCGAAATCTCTCTCTTTGTCCACAGAATAAAAATCTATTTCTTGTG
TTTGTGTTGTATTTACTTAAAAAAATAAAGGAGATCAGTTTGGGAGAAAGCAGC
AAAGAAGAAAA

AZprom_17 [SEQ ID NO:17]
TGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGA
CAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATGGCTTATGAAT
CTGACGTGGAACAAGAAGAAAGCGTATCTGCCACGTGGCACAATACAACACCT
CAACAGCCTTATCTGGACATCAACAGAGAGATCCTCCAATATCCGAAACCAAC
CAATCAGCTTCACTCTTGGTTTCACCTCCATACATTAACACACCAACAAGGCTT
ATGAATCTGACGTGGAACAAGAAGAAAGCGTATCTGCCACGTGGCACAATACA
ACACCTCAACAGCCTTATCTGGACATCAACAGAGAGATCCTCCAATATCCGAA
ACCAACCAATCAGCTTCACTCTTGGTTTCACCTCCATACATTAACACACCAACA
ATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCC
TTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGAcacgctgCATATTGGAAG
TTAAAGGAAAAGAGAGAAAGAGAAATCTTTCTGTCTAAGTGTAATTAACA

AZprom_18 [SEQ ID NO:18]
GCTGGTGGACTAGAGATTGCCACGTAAGACTACTAAACGATAAAACAAAAATC
TTAAAATCCAATGAATGAACAGATAAAGATTACTTCAGATATAACAAACGTTACA
ATATCCCGCTGGTGGACTAGAGATTGCCACGTAAGACTACTAAACGATAAAAC
AAAAATCTTAAAATCCAATGAATGAACAGATAAAGATTACTTCAGATATAACAAA
CGTTACAATATCCCGTAAAACGTCAAAGTTTTAAGGTGATATCACATAATTCAT
GAGATTATTTTGGTTATTGGATAAAGTGACGATGAAATCGATTCATAATTGAAAT
TAAACTGGTAGTGATGATGAACTAATTACCATTGAAAATATTGATATTTATGTGT
GGGTAAATGCTTTTTATAATGTTACTTTATATTAATGTTTCGATCATCAGAATCT
ATATTTTCAAAATGTTATACTTTAAGTTTTAGTTATTGGGTTGTAGCAAAAATCAT
TCTTGTCACGAGGGTGTAAGTAAGTGTAACCGTTGAAGTATTCAGTGGCTCAT
AACTTGTGGTCACAAAACGCTTGGCTGCAATGAAAAAATCAAAACAAATGCTG
GTGGACTAGAGATTGCCACGTAAGACTACTAAACGATAAAACAAAAATCTTAAA
ATCCAATGAATGAACAGATAAAGATTACTTCAGATATAACAAACGTTACAATATC
CCTATATAATCCAACACTATCGAACCAGTTTTAATCACTCTCACTCACAAGTTA
GTCATAAAAAAAAAAAAAACACAAAAAAGTTTCA

AZprom_19 [SEQ ID NO:19]
CAATCCGTCATATTTTTTCAACCTAAATGCAACGTCATCAAAAAGTAGACAGCG
TAGCATAACATGAAATCTTGAAAACATTGCGTGAGAAGGAAGAATTAGGATAAG
GAACTGGGAAAGTGTTGTTTATGAACTTGAATAAATGTTCCCATTAATCATGGA
TCAAATGTGTCACCATGCAAATGTGTGATCTTGAAATTCTGCCACGAAAAGCTT
TAGACAGGACGTGGCATGCCACGTCAACAAACAAATAGATCCAAAAATGGTTG
AGCTGTTTTGGATTTGGCAATTAATTGCATCGTGGCAAATTTTTTCACTCTCTCA
TCGTCATTATCGTCATCATCACTCTCTCTCTCTACGCTTGCACAATCCGTCATA
TTTTTTCAACCTAAATGCAACGTCATCAAAAAGTAGACAGCGTAGCATAACATG
AAATCTTGAAAACATTGCGTGAGAAGGAAGAATTAGGATAAGGAACTGGGAAA
GTGTTGTTTATGAACTTGAATAAATG
TTCCCATTAATCATGGATCAAATGTGTCACCATGCAAATGTGTGATCTTGAAAT
TCTGCCACGAAAAGCTTTAGACAGGACGTGGCATGCCACGTCAACAAACAAAT

AGATCCAAAAATGGTTGAGCTGTTTTGGATTTGGCAATTAATTGCATCGTGGCA
AATTTTTTCACTCTCTCATCGTCATTATCGTCATCATCACTCTCTCTCTCTACGC
TTGCAGTCCGTCAGATATAGGAAATATGTAAAACCTTATCATTATATATAGGGT
GGTGGGCAACTATGCAATGACCATATTGGAAGTTAAAGGAAAAGAGAGAAAGA
GAAATCTTTCTGTCTAAGTGTAATTAACA

AZprom_20 [SEQ ID NO:20]
AAAGGAAATAGTGGTGTTGCATCAAGTTATGGACAATATAAGGAAGCAAACAG
TACTCTAGCTATCAAATTAGTTTCCACTTCTAAACCATAAAGCCAAGGAAAAGG
GACTCTAAAAAAATTAACCAACGATAAGATTACTAGGTCTTACACGTGGCACCT
CCATTGTGGTGACTAAATGAAGAGTGGCTTAGCTCAAAATATAATTTTCCAACC
TTTCATGTGTGGATATTAAGTTTTGTGTAGTGAATCAAGAACCACATAATCCAAT
GGTTAGCTTTATTCCAAGATGAGGGGGTTGTTGATTTTTGTCCGTCAGATATAG
GAAATATGTAAAACCTTATCATAAAGGAAATAGTGGTGTTGCATCAAGTTATGG
ACAATATAAGGAAGCAAACAGTACTCTAGCTATCAAATTAGTTTCCACTTCTAA
ACCATAAAGCCAAGGAAAAGGGACTCTAAAAAAATTAACCAACGATAAGATTAC
TAGGTCTTACACGTGGCACCTCCATTGTGGTGACTAAATGAAGAGTGGCTTAG
CTCAAAATATAATTTTCCAACCTTTCATGTGTGGATATTAAGTTTTGTGTAGTGA
ATCAAGAACCACATAATCCAATGGTTAGCTTTATTCCAAGATGAGGGGGTTGTT
GATTTTTGTCCGTCAGATATAGGAAATATGTAAAACCTTATCATTATTTTGTAAA
CGTTTTATAATAATAGAATTATTTTTTCACTCTCTCATCGTCATTATCGTCATCAT
CACTCTCTCTCTCTACGCTTGCATATATATAAACCATTGCAGAGCCGCGGAGT
GTCAAGCATCGTGGTGGAGTAGAGTGAGAGAACCGAAGCCAAAGGCAACTCC
ATTATTCTCTCTCGTGTTCTCGTGATATTGGTTTTCCGGCGCCGGAGCTTGTCG
GTCCGTCCGTCCGTCC

AZprom_21 [SEQ ID NO:21]
TGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGA
CAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATTACATACTTTCC
AAATTCATAGAAAATCTTCAATGTTGTATTAGATTATGTCTTTAACACACAAAAT
TGACGCATTAACGCCTCTATGTAGAGAATATTATTGTACTTAATTTCCATTTCTC
AATATATATGCAACATATTCACACACACACACACATATATATATACATATATACA
TACATACATATATATATATATATATATATATGTGTGTGTGTGTGTGTATATATATA
TATACATATATATATATATATACATATACATATATATATATATATATATATATAT
ATGTATCATGACAAAAAGAAAATGTTCGATTCATCTTTTTATGGAGGATTTTGTT
TGTTTCAGTTTTTTGTTTGTTTTAGTTTTTTGTTTGTTTTAGTGAGTAAGTACAGC
TCTCATAAATTCACATGATTTCAACCATAGAAATATAGTCTCAATATTGTCCACG
TATAAGCTCATTATCCTTTCCGTGGATTTGTTTTGTTTTTCTATTGTGTAACGAG
AAATATTCGAATAAAAACATAGATGCAATACTCCTGTCCTGGTGAGTCCTTTTTT
CATGTTCTAGACGGTGTTAATGATGGATGATGTTAAATGACATCGTTTTAATAC
TAATTGTTTTTTAATTTACAAAACTCTCAACAAATGATTAGTTGGGTTAGTTATTC
ATAGGAAAGCGGACGAGCATGTCGTTATAATTAAAAAAATATCAAAAGAGTAAA
CAAAAAAGGAAAAAGACTAATTATTTAGATAATAATAATATCCACAAAAATATTC

GAATTCTTCAATCCTGAGTTTGCTCTGTGGATGAGTTTCTGTATCATTGATACTT
GATACCTGTAATTCACACACCTCATATCTCATACTTCATCTATAAATACCCAATT
CATTTTGCTCAAAGTCTCAACACTGAGCATACCCAATATTCAGGTGATCTAATT
TAACGTTTGCATGAGTATTTTCTTAATAAAATTTATGTTGGGTTTACAGTATCTA
TTGGGTGGATTTCTTAAACGGATTGTGGTTTGATTAATAAAAAATCTTAATGAG
AAGTTTGTGATAATATGCTGAAATGGGTTGTTTTTGTGTTAATTTTTCAGGGTTG
GAGGGGAATTAAGTATTAAGCAAGGGTGTGAGTA

C1 [SEQ ID NO:22]
TGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTAC
AAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGA
CAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGA
CGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATAT
AAGGAAGTTCATTTCATTTGGAGAGGACTAAACC

C2 [SEQ ID NO:23]
CAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAG
TCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGA
AACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTA
GAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTAT
CGTTCA
AGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGC
ATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATG
TGATAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGA
ATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAA
GGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTC
ATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGA
TAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATG
GACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTC
TTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCAC
AATCCCACTATCCTTCGCAAGACCTTCCTCTATATAAGGAAGTTCAT

TTCATTTGGAGAGGACACGCTGAAATCACCAGTCTCTCTCTACAAATCTATCTC
TCTCGAGCTTTCGCAGAACTGTCGATCGACC

VPP_mod [SEQ ID NO: 24]
CAATCCGTCATATTTTTTCAACCTAAATGCAACGTCATCAAAAAGTAGACAGCG
TAGCATAACATGAAATCTTGAAAACATTGCGTGAGAAGGAAGAATTAGGATAAG
GAACTGGGAAAGTGTTGTTTATGAACTTGAATAAATGTTCCCATTAATCATGGA
TCAAATGTGTCACCATGCAAATGTGTGATCTTGAAATTCTGCCACGAAAAGCTT
TAGACAGGACGTGGCATGCCACGTCAACAAACAAATAGATCCAAAAATGGTTG
AGCTGTTTTGGATTTGGCAATTAATTGCATCGTGGCAAATTTTTTCACTCTCTCA
TCGTCATTATCGTCATCATCACTCTCTCTCTCTACGCTTGCA

RBCss_mod [SEQ ID NO: 25]
AAAGGAAATAGTGGTGTTGCATCAAGTTATGGACAATATAAGGAAGCAAACAG
TACTCTAGCTATCAAATTAGTTTCCACTTCTAAACCATAAAGCCAAGGAAAAGG
GACTCTAAAAAAATTAACCAACGATAAGATTACTAGGTCTTACACGTGGCACCT
CCATTGTGGTGACTAAATGAAGAGTGGCTTAGCTCAAAATATAATTTTCCAACC
TTTCATGTGTGGATATTAAGTTTTGTGTAGTGAATCAAGAACCACATAATCCAAT
GGTTAGCTTTATTCCAAGATGAGGGGGTTGTTGATTTTTGTCCGTCAGATATAG
GAAATATGTAAAACCTTATCAT

inverted DR region of Ft-PEPc promoter [SEQ ID NO: 26]
TACATACTTTCCAAATTCATAGAAAATCTTCAATGTTGTATTAGATTATGTCTTTA
ACACACAAAATTGACGCATTAACGCCTCTATGTAGAGAATATTATTGTACTTAAT
TTCCATTTCTCAATATATATGCAACATATTCACACACACACACACATATATATAT
ACATATATACATACATACATATATATATATATATATATATATATGTGTGTGTGTGTGT
GTATATATATATATACATATATATATATATATACATATACATATATATATATATATA
TATATATATATATGTATCATGACAAAAAGAAAATGTTCGATTCATCTTTTTATGG
AGGATTTTGTTTGTTTCAGTTTTTTGTTTGTTTTAGTTTTTTGTTTGTTTTAGTGA
GTAAGTACAGCTCTCATAAATTCACATGATTTCAACCATAGAAATATAGTCTCAA
TATTGTCCACGTATAAGCTCATTATCCTTTCCGTGGATTTGTTTTGTTTTTCTAT
TGTGTAACGAGAAATATTCGAATAAAAACATAGATGCAATACTCCTGTCCTGGT
GAGTCCTTTTTTCATGTTCTAGA

YPet_frw [SEQ ID NO: 27]
CTCAGTAAGTGGGGAAGGTGAAGGC

YPet_rev [SEQ ID NO: 28]
TGCCAGCTGAACACCTCCATCCTCG

cDNA sequence encoding the mature SP-B peptide [SEQ ID NO: 29]
TTCCCCATTCCTCTCCCCTATTGCTGGCTCTGCAGGGCTCTGATCAAGCGGAT
CCAAGCCATGATTCCCAAGGGTGCGCTAGCTGTGGCAGTGGCCCAGGTGTGC
CGCGTGGTACCTCTGGTGGCGGGCGGCATCTGCCAGTGCCTGGCTGAGCGC
TACTCCGTCATCCTGCTCGACACGCTGCTGGGCCGCATGCTGCCCCAGCTGG
TCTGCCGCCTCGTCCTCCGGTGCTCCATG

AZ18_1xSP-B (YPet – linker1 – TEV-1xSP-B – TEV – linker2 - YPet)[SEQ ID NO: 30]

ATGGCTTCTAAAGGAGAAGAATTGTTTACTGGTGTAGTGCCTATCCTTGTTGAA
TTGGATGGTGATGTTAATGGCCACAAATTCTCAGTAAGTGGGGAAGGTGAAGG
CGACGCTACATATGGAAAGCTCACTCTTAAGCTGCTATGTACTACTGGAAAGTT
GCCCGTCCCTTGGCCAACATTGGTTACAACACTTGGATATGGGGTTCAGTGTT
TTGCAAGATATCCAGACCACATGAAGCAACATGACTTCTTCAAATCTGCTATGC
CAGAGGGATATGTCCAAGAAAGGACAATATTCTTCAAGGATGATGGAAATTAC
AAGACAAGAGCTGAGGTGAAGTTTGAAGGCGATACACTGGTTAACAGAATCGA
ACTAAAGGGCATTGATTTCAAGGAAGATGGAAATATCTTAGGGCATAAGCTGG
AGTATAACTATAACTCTCATAATGTTTACATAACCGCAGACAAGCAGAAAAACG
GGATTAAGGCAAATTTCAAGATTAGGCATAACATCGAGGATGGAGGTGTTCAG
CTGGCAGACCATTATCAGCAAAATACACCAATAGGTGATGGCCCAGTCTTATT
GCCTGATAACCATTATCTCTCCTATCAGAGTGCTCTCTTTAAGGACCCTAACGA
AAAGAGAGATCATATGGTCTTACTTGAGTTCTTGACTGCAGCCGGAATTACCG
AAGGGATGAACGAACTTTACAAGGAGCCCGAGCCGGAGCCTGAGCCAGAACC
AGAGCCAGAGCCGCCCGAAAATCTTTACTTTCAAGGTTTCCCTATCCCACTGC
CTTACTGCTGGCTTTGTCGAGCCCTTATCAAGAGGATTCAAGCAATGATCCCA
AAAGGTGCCTTAGCAGTCGCCGTGGCCCAAGTTTGCAGAGTAGTGCCTCTTGT
TGCTGGCGGAATTTGCCAATGTCTTGCAGAGCGTTACAGCGTTATTCTGTTGG
ATACTCTGTTGGGCAGAATGCTACCTCAATTGGTGTGCCGTCTGGTGTTACGT
TGTAGCATGCCAGAAAATTTGTATTTTCAAGGGGAACCTGAGCCTGAACCTGA
ACCGGAGCCCGAACCAGAACCAATGTCAAAAGGAGAGGAACTTTTTACTGGA
GTGGTTCCAATCCTCGTGGAACTTGACGGCGATGTTAACGGTCACAAGTTTAG
TGTTTCTGGAGAAGGAGAAGGGGATGCTACGTACGGAAAACTTACACTCAAGT
TGTTGTGTACTACGGGAAAACTTCCAGTACCCTGGCCAACTCTAGTCACCACT
CTCGGTTATGGTGTTCAGTGCTTTGCTCGGTATCCTGATCATATGAAACAACAC
GACTTTTTCAAATCGGCAATGCCGGAGGGTTATGTTCAGGAGAGGACCATATT
TTTCAAAGATGATGGGAATTACAAAACCCGAGCTGAAGTCAAATTCGAAGGAG
ATACTCTCGTAAATCGCATTGAGTTAAAGGGTATTGACTTTAAAGAGGACGGCA
ACATACTAGGCCATAAACTTGAGTACAATTACAATTCACATAATGTGTATATAAC
TGCTGACAAACAAAAGAATGGTATTAAGGCTAATTTCAAAATTCGGCACAATAT
TGAAGATGGTGGTGTACAGTTGGCGGATCATTACCAGCAAAACACACCTATAG
GAGATGGTCCTGTTCTATTACCTGATAATCACTATCTCTCTTATCAATCCGCATT
GTTTAAAGACCCAAACGAGAAAAGGGATCACATGGTGCTTTTGGAGTTTTTGA
CTGCTGCTGGTATTACCGAGGGGATGAATGAGTTATACAAAGCAGATCTCCAA
AAGGACGAACTTTAA

AZ18_2xSP-B (YPet – linker1 – TEV-2xSP-B – TEV – linker2 – Ypet) [SEQ ID NO: 31]

ATGGCTTCTAAAGGAGAAGAATTGTTTACTGGTGTAGTGCCTATCCTTGTTGAA
TTGGATGGTGATGTTAATGGCCACAAATTCTCAGTAAGTGGGGAAGGTGAAGG
CGACGCTACATATGGAAAGCTCACTCTTAAGCTGCTATGTACTACTGGAAAGTT
GCCCGTCCCTTGGCCAACATTGGTTACAACACTTGGATATGGGGTTCAGTGTT
TTGCAAGATATCCAGACCACATGAAGCAACATGACTTCTTCAAATCTGCTATGC

CAGAGGGATATGTCCAAGAAAGGACAATATTCTTCAAGGATGATGGAAATTAC
AAGACAAGAGCTGAGGTGAAGTTTGAAGGCGATACACTGGTTAACAGAATCGA
ACTAAAGGGCATTGATTTCAAGGAAGATGGAAATATCTTAGGGCATAAGCTGG
AGTATAACTATAACTCTCATAATGTTTACATAACCGCAGACAAGCAGAAAAACG
GGATTAAGGCAAATTTCAAGATTAGGCATAACATCGAGGATGGAGGTGTTCAG
CTGGCAGACCATTATCAGCAAAATACACCAATAGGTGATGGCCCAGTCTTATT
GCCTGATAACCATTATCTCTCCTATCAGAGTGCTCTCTTTAAGGACCCTAACGA
AAAGAGAGATCATATGGTCTTACTTGAGTTCTTGACTGCAGCCGGAATTACCG
AAGGGATGAACGAACTTTACAAGGAGCCCGAGCCGGAGCCTGAGCCAGAACC
AGAGCCAGAGCCGCCCGAAAATCTTTACTTTCAAGGTTTCCCTATCCCACTGC
CTTACTGCTGGCTTTGTCGAGCCCTTATCAAGAGGATTCAAGCAATGATCCCA
AAAGGTGCCTTAGCAGTCGCCGTGGCCCAAGTTTGCAGAGTAGTGCCTCTTGT
TGCTGGCGGAATTTGCCAATGTCTTGCAGAGCGTTACAGCGTTATTCTGTTGG
ATACTCTGTTGGGCAGAATGCTACCTCAATTGGTGTGCCGTCTGGTGTTACGT
TGTAGCATGTTTCCTATTCCTCTACCCTACTGTTGGCTTTGCAGAGCCCTCATC
AAAAGGATTCAGGCTATGATACCAAAAGGAGCTTTAGCTGTAGCAGTAGCTCA
AGTGTGTAGAGTCGTCCCCCTTGTTGCAGGGGGAATTTGTCAGTGCCTTGCTG
AGCGTTATAGTGTTATCTTGCTCGATACTTTACTTGGTCGTATGTTGCCCCAAC
TCGTATGCCGATTAGTTTTGAGGTGTTCTATGCCAGAAAATTTGTATTTTCAAG
GGGAACCTGAGCCTGAACCTGAACCGGAGCCCGAACCAGAACCAATGTCAAA
AGGAGAGGAACTTTTTACTGGAGTGGTTCCAATCCTCGTGGAACTTGACGGCG
ATGTTAACGGTCACAAGTTTAGTGTTTCTGGAGAAGGAGAAGGGGATGCTACG
TACGGAAAACTTACACTCAAGTTGTTGTGTACTACGGGAAAACTTCCAGTACCC
TGGCCAACTCTAGTCACCACTCTCGGTTATGGTGTTCAGTGCTTTGCTCGGTA
TCCTGATCATATGAAACAACACGACTTTTTCAAATCGGCAATGCCGGAGGGTT
ATGTTCAGGAGAGGACCATATTTTTCAAAGATGATGGGAATTACAAAACCCGA
GCTGAAGTCAAATTCGAAGGAGATACTCTCGTAAATCGCATTGAGTTAAAGGG
TATTGACTTTAAAGAGGACGGCAACATACTAGGCCATAAACTTGAGTACAATTA
CAATTCACATAATGTGTATATAACTGCTGACAAACAAAAGAATGGTATTAAGGC
TAATTTCAAAATTCGGCACAATATTGAAGATGGTGGTGTACAGTTGGCGGATCA
TTACCAGCAAAACACACCTATAGGAGATGGTCCTGTTCTATTACCTGATAATCA
CTATCTCTCTTATCAATCCGCATTGTTTAAAGACCCAAACGAGAAAAGGGATCA
CATGGTGCTTTTGGAGTTTTTGACTGCTGCTGGTATTACCGAGGGGATGAATG
AGTTATACAAAGCAGATCTCCAAAAGGACGAACTTTAA

Linker 1[SEQ ID NO: 32]
GAGCCCGAGCCGGAGCCTGAGCCAGAACCAGAGCCAGAGCCG

Linker 2[SEQ ID NO: 33]
GAACCTGAGCCTGAACCTGAACCGGAGCCCGAACCAGAACCA

FIG. 4

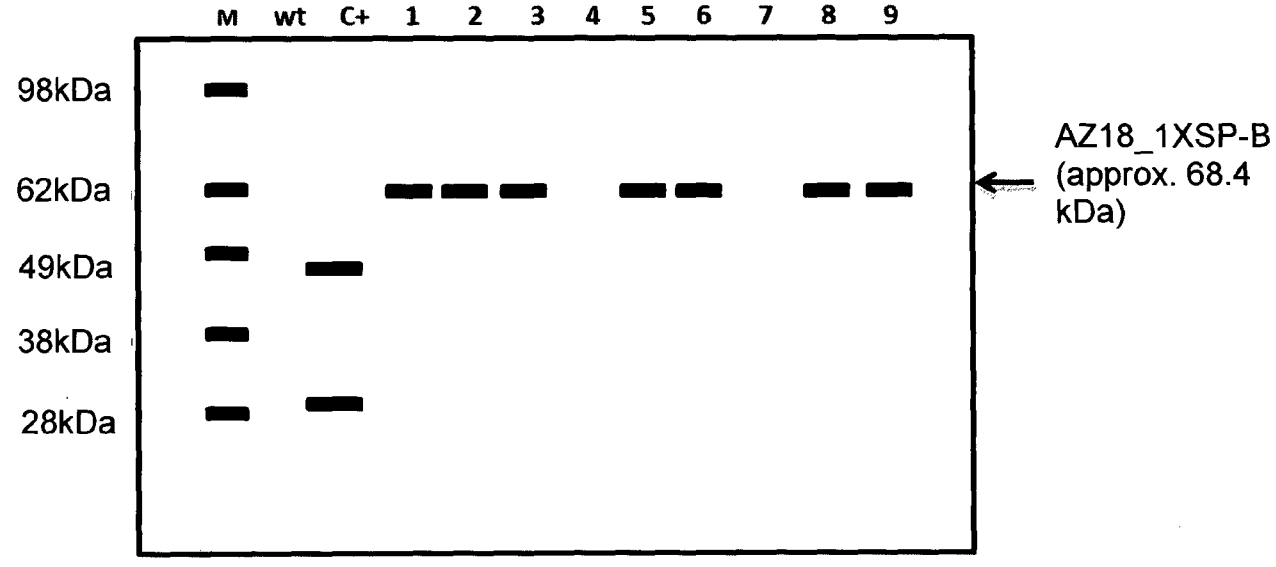

M
wt
C+
1
2
3
4
5
6
7
8
9
98kDa
62kDa
49kDa
38kDa
28kDa
AZ18_1XSP-B
(approx. 68.4
kDa)

FIG. 5

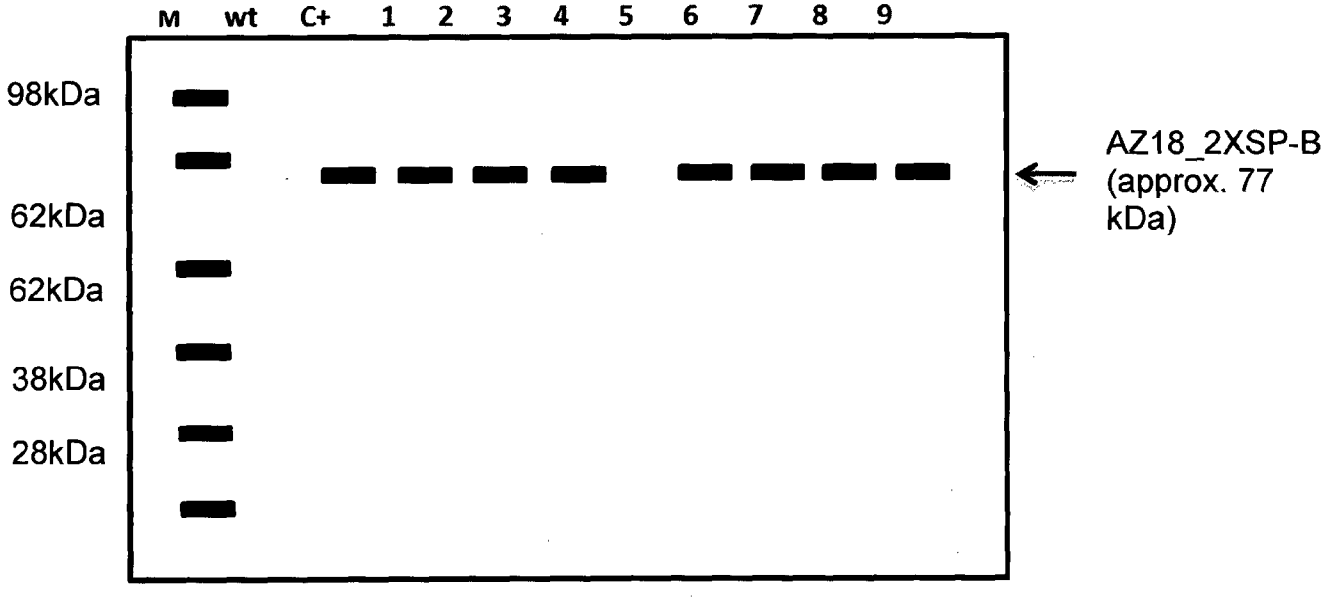

M
wt
C+
1
2
3
4
5
6
7
8
9
98kDa
62kDa
62kDa
38kDa
28kDa
AZ18_2XSP-B
(approx. 77
kDa)

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2015/052503

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/82
ADD. A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, CHEM ABS Data, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2012/038717 A1 (CAMBRIDGE ADVANCED TECH [GB]; DAVENPORT SUSAN [GB]; LE LAY PASCALINE [) 29 March 2012 (2012-03-29)	1-4,6,9, 12, 15-17, 20,21
Y	page 10, line 9 - page 15, line 24; claims; sequence 2 & DATABASE Geneseq [Online] 24 May 2012 (2012-05-24), "Pisum sativum plastocyanin promoter DNA SEQ:2.", retrieved from EBI accession no. GSN:AZU50004 Database accession no. AZU50004 sequence	5,7,8, 10,11, 13,14, 18,19, 22-24

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 July 2015	30/10/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Burkhardt, Peter

1

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2015/052503

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00/56906 A1 (MERISTEM THERAPEUTICS [FR]; RANCE IANN [FR]; GRUBER VERONIQUE [FR]; TH) 28 September 2000 (2000-09-28)	1-4,6,9, 12, 15-17, 20,21
Y	page 9, line 12 - page 11, line 27; claims; sequence 1 & DATABASE Geneseq [Online] 8 February 2001 (2001-02-08), "Nucleotide sequence of the petE promoter from plastocyanin gene.", retrieved from EBI accession no. GSN:AAA96457 Database accession no. AAA96457 sequence	5,7,8, 10,11, 13,14, 18,19, 22-24

Y	VENTER ET AL: "Synthetic promoters: genetic control through cis engineering", TRENDS IN PLANT SCIENCE, ELSEVIER SCIENCE, OXFORD, GB, vol. 12, no. 3, 16 March 2007 (2007-03-16) , pages 118-124, XP005920450, ISSN: 1360-1385, DOI: 10.1016/J.TPLANTS.2007.01.002 cited in the application page 119, column 2, paragraph 1; figure 1	5,7,8, 10,11, 13,14, 18,19, 22-24

Y	POTENZA C ET AL: "INVITED REVIEW: TARGETING TRANSGENE EXPRESSION IN RESEARCH, AGRICULTURAL, AND ENVIRONMENTAL APPLICATIONS: PROMOTERS USED IN PLANT TRANSFORMATION", IN VITRO CELLULAR & DEVELOPMENT BIOLOGY. PLANT, GAITHERSBURG, MD, US, vol. 40, no. 1, 1 January 2004 (2004-01-01), pages 1-22, XP009045613, ISSN: 1054-5476, DOI: 10.1079/IVP2003477 page 15, column 1, paragraph 2 - column 2, paragraph 2	5,7,8, 10,11, 13,14, 18,19, 22-24

Y	WO 03/020899 A2 (DIADEXUS INC [US]; SUN YONGMING [US]; LIU CHENGHUA [US]; CHEN SEI-YU [) 13 March 2003 (2003-03-13) claims	5,7,8, 10,11, 13,14, 18,19, 22-24

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2015/052503

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-24(partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-24(partially)

   relating to the promoter of SEQ ID NO:1 and products and methods comprising said promoter

   ---

2. claims: 1-24(partially)

   relating to the promoter of SEQ ID NO:3 and products and methods comprising said promoter

   ---

3. claims: 1-24(partially)

   relating to the promoter of SEQ ID NO:4 and products and methods comprising said promoter

   ---

4. claims: 1-24(partially)

   relating to the promoter of SEQ ID NO:18 and products and methods comprising said promoter

   ---

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2015/052503

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012038717 A1	29-03-2012	AR 083085 A1	30-01-2013
		CA 2812380 A1	29-03-2012
		CN 103403021 A	20-11-2013
		EP 2619220 A1	31-07-2013
		JP 2013543377 A	05-12-2013
		US 2013206155 A1	15-08-2013
		WO 2012038717 A1	29-03-2012
WO 0056906 A1	28-09-2000	AU 767285 B2	06-11-2003
		AU 3185300 A	09-10-2000
		CA 2366228 A1	28-09-2000
		CN 1351669 A	29-05-2002
		EP 1163351 A1	19-12-2001
		FR 2791358 A1	29-09-2000
		JP 2002539795 A	26-11-2002
		WO 0056906 A1	28-09-2000
WO 03020899 A2	13-03-2003	AU 2002331780 A1	18-03-2003
		WO 03020899 A2	13-03-2003