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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 July 2008 (10.07.2008)

PCT

(10) International Publication Number
WO 2008/080954 A1

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

(21) International Application Number: PCT/EP2007/064590

(22) International Filing Date: 27 December 2007 (27.12.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
UD2006A000280 29 December 2006 (29.12.2006) IT

(71) Applicant *(for all designated States except US)*: **UNIVERSITA' DEGLI STUDI DI UDINE** [IT/IT]; Via Palladio, 8, I-33100 Udine (IT).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MARCHETTI, Stefano** [IT/IT]; Via dei Brazzà, 29, I-33010 Pagnacco (IT). **DE AMICIS, Francesca** [IT/IT]; Via Pradamano, 5/A, I-33042 Buttrio (IT). **PATTI, Tamara** [IT/IT]; Via Beorcjte, 15, I-33030 Buja (IT).

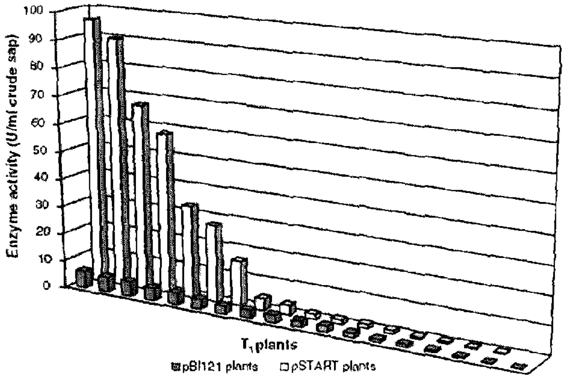
(74) Agents: PETRAZ, Gilberto et al.; GLP SRL, Piazzale Cavedalis, 6/2, I-33100 Udine (IT).

(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, BR, BW, BY, BZ, CA, CH, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
- *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(54) Title: ARTIFICIAL DNA SEQUENCE WITH OPTIMIZED LEADER FUNCTION IN 5' (5'-UTR) FOR THE IMPROVED EXPRESSION OF HETEROLOGOUS PROTEINS IN PLANTS



(57) Abstract: A nucleotide leader sequence 5'-UTR comprises elements favorable to gene expression, such as repeated CAA trinucleotide elements in combination with repeated CT dinucleotide elements.

WO 2008/080954 A1

"ARTIFICIAL DNA SEQUENCE WITH OPTIMIZED LEADER FUNCTION IN 5' (5'-UTR) FOR THE IMPROVED EXPRESSION OF HETEROLOGOUS PROTEINS IN PLANTS"

* * * * *

FIELD OF THE INVENTION

The present invention concerns an artificial DNA sequence for improving the expression of heterologous proteins in plants.

BACKGROUND OF THE INVENTION

In the field of biotechnology, there is a strongly felt need to enhance the level of expression of genes introduced into the relative organisms. This level is often unsatisfactory, and represents a barrier to the industrial application of innovations in plant and animal biotechnology. There is a quantity of data to support the importance of the leader region in regulating the levels of gene expression, while there are various structural elements that characterize the regulation capacity thereof.

In this case, the untranslated leader sequence in 5' (5'-UTR), as it is proposed in widely diffused vectors (e.g. pBI121 and derivatives, pCAMBIA and derivatives), has numerous defects which make it unsuitable to direct adequate levels of gene expression in genetically modified organisms. In particular, when yields are to be maximized (e.g. the use of plants as cellular factories for compounds useful to man), it is necessary to eliminate the production constraints exerted by the 5'-UTR sequence. To this end, the leader Ω (a sequence that exists naturally in tobacco mosaic virus, TMV) has been proposed in plants. However, this too has imperfections and redundancies that render it open to improvement.

It is known that the region poly(CAA) in the translational enhancer present in the leader Ω of TMV (Gallie and Walbot 1992 Nucleic Acids Res., 20, 4631-4638) significantly enhances the expression levels, that is, it has a positive effect on the translation levels of heterologous proteins in vitro and in vivo (Gallie et al. 1988a Nucleic Acids Res., 16, 883-893, Gallie et al. 1988b Nucleic Acids Res., 16, 8675-8694, Gallie 2002 Nucleic Acids Res., 30, 3401-3411). In the leader Ω, a poly(CAA) sequence is associated with 3 repeats of the sequence ACAATTAC (Gallie et al. 1988a), but delection studies have shown that the regulator element responsible for enhancing the expression levels may consist of a single copy of the sequence ACAATTAC in combination with the motif (CAA)n (Gallie and Walbot 1992).

It is also known that the transcription initiation site (*Inr*) of the CaMV 35S promoter (Guilley et al. 1982 Cell, 30, 763-773) favours an efficient capping of the mRNA.

Furthermore, it is known that many plant leaders (Bolle et al. 1996 Plant Mol. Biol. 32, 861-868) have a sequence rich in CT elements and that the CT-rich sequences can alter the transcription levels (Chen et al. 1996 J. Virol., 70, 8411-8421).

It is also known that sequences which have a length of more than 40 nucleotides promote the recognition of the first AUG as authentic initial translation codon (Kozak 1989 J. Cell. Biol., 108, 229-241). For example, it has been observed that the extension of the leader from 29 to 74 nt causes an increase in the translation level of mRNA in vitro (Kozak 1991, J. Biol. Chem., 266, 19867-19870) and in vivo (Gallie and Walbot 1992). Leader sequences with a greater content of A/T cause higher expression levels since the formation of segments of double strand mRNA, due to the folding of the molecule over itself, is discouraged. In fact, it is certain that such secondary structures have a depressing effect on the translation efficiency (Pelletier and Sonenberg 1985 Cell, 40, 515-526; Kozak 1986 Proc. Natl. Acad. Sci. USA, 83, 2850-2854). Moreover, it has been noticed that the introduction of portions of 5'-UTRs of viral origin into plant leaders can be reflected in an increase in the level of expression of reporter proteins (Dowson Day et al. 1993 Plant Mol. Biol., 23, 97-109).

Purpose of the present invention is therefore to obviate the shortcomings of the state of the art and to achieve a leader sequence that increases the expression levels of recombinant proteins in plants.

The Applicant has devised, tested and embodied the present invention to overcome the shortcomings of the state of the art and to obtain these and other purposes and advantages.

SUMMARY OF THE INVENTION

The present invention is set forth and characterized in the independent claims, while the dependent claims describe other characteristics of the invention or variants to the main inventive idea.

In accordance with the above purpose, an artificial DNA sequence having a leader function in 5' (5'-UTR), hereafter indicated by LL-TCK, according to the present invention simultaneously comprises elements favorable to gene expression, such as repeated trinucleotide elements CAA in combination with repeated dinucleotide elements CT.

The LL-TCK sequence according to the present invention was obtained by means of artificial synthesis and is the fruit of the intellect, since it does not exist in nature.

According to an advantageous solution, the LL-TCK sequence according to the present invention provides the combination of trinucleotide elements CAA with dinucleotide elements CT and a modification of the sequences that activate translation present in the leader Ω.

According to a variant, the sequence according to the present invention contains a poly(CAA) region, that is, an oligonucleotide consisting of 2 or more copies of the CAA element, preferably but not necessarily contiguous with each other.

According to another variant, the sequence according to the present invention contains a poly(CT) region, that is, an oligonucleotide consisting of 2 or more copies of the CT element, preferably but not necessarily contiguous with each other.

A variant of the present invention provides that the sequence contains one or more copies of the octamer ACAATTAC.

A sequence obtained from the combination of the sequences with a poly(CAA) region and those with a poly(CT) region also comes within the field of the present invention.

A sequence obtained from the combination of the sequences with a poly(CAA) region and those with one or more copies of the octamer ACAATTACC also comes within the field of the present invention.

Furthermore, a sequence obtained from the combination of the sequences with a poly(CT) region and those with one or more copies of the octamer ACAATTACC also comes within the field of the present invention.

According to the present invention, it is possible to provide a sequence obtained from the combination of the sequences with a poly(CAA) region, those with a poly(CT) region and those with one or more copies of the octamer ACAATTACC.

Furthermore, again according to the present invention, it is possible to provide a sequence obtained from the combination of one or more of the above sequences with the CaMV 35S *Inr* site, that is, the transcription initiation site of the cauliflower mosaic virus 35S promoter.

The LL-TCK sequence according to the present invention is thus able to increase the expression levels of heterologous proteins in transgenic plants. According to an advantageous solution of the present invention, the new sequence LL-TCK was synthesized so as to create a combination of the following elements according to an original pattern, unique of its kind:

(1) transcription start site (*Inr*) of the CaMV 35S promoter for an efficient mRNA capping;

(2) poly(CAA) region similar to the translational enhancer present in the TMV leader Ω;

(3) a sequence rich in CT elements, like many plant leaders.

Furthermore, the LL-TCK sequence has a length of more than 40 nucleotides in order to promote the recognition of the first AUG as the authentic translation start codon (Kozak 1989) and an overall content of G+C of less than 40%.

According to a particular solution of the present invention, the LL-TCK sequence is the one shown in SEQ ID NO: 1 (5'-3').

It is possible to foresee that small mutations in the LL-TCK sequence do not alter its effectiveness and for this reason the present invention also refers to leader sequences derived from the present sequence, for example following deletion or duplication of a CAA triplet, substitution or deletion of a single base, etc.

The innovation of LL-TCK consists in the fact that it joines in a single leader a modified poly(CAA) element, an octamer from the TMV leader Ω and a CT-rich sequence of plant origin.

Therefore, the artificial sequence LL-TCK according to an advantageous solution of the present invention provides the presence of a single octamer ACAATTAC associated with 9 CAA repetitions located in position 5' with respect to the octamer.

Since the triplet ATT inside the element ACAATTAC can represent a non-canonical translation start site (Tyc et al. 1984 Eur. J. Biochem., 140, 503-511, Schmitz et al. 1996 Nucleic Acids Res., 24, 257-263), in the LL-TCK leader this triplet has been put in frame with a stop codon.

Furthermore, in the artificial LL-TCK leader an element $(CT)_4$ has been added to the 3' end of the regulator element obtained from the union of the octamer ACAATTAC with the poly(CAA) sequence. The combination of these two elements, for each of which the positive effect on gene expression is known, has never been found in nature or previously made by man.

The LL-TCK leader, combining these two elements, causes an enhancement of both the translation level and the transcription level of the gene concerned.

This effect has been demonstrated by comparing the expression levels of the gene *uid*A obtained in tobacco plants (*Nicotiana tabaccum*) transformed with the constructs 35S-LL-TCK:: *uid*A (pSTART) and 35S::*uid*A (pBI121 with original leader). The vector pSTART was obtained by replacing the leader sequence in pBI121 (Clontech) with LL-TCK. In particular, the object of the replacement and manipulation was the pBI121 sequence comprised between the region *Inr* (ACACG) and the restriction site *Xba* I (TCTAGA). The nucleotide sequence flanking the initial translation codon of *uid*A was kept unvaried in the two constructs so as to prevent variability in the codon AUG recognition.

The choice of using the enzyme beta-glucuronidase (GUS) encoded by the gene *uid*A as reporter protein was determined by the fact that in tobacco no native GUS-like activity can be observed, and the level of expression of the transgene *uid*A can be measured by means of a fluorimetric test (Jefferson et al. 1987 EMBO J., 6, 3901-3907) characterized by considerable sensitivity, accuracy, speed and ease of execution.

The fluorimetric readings relating to the enzymatic GUS activity, measured as described by Jefferson (1987) in plants regenerated after transformation (generation T_1) have shown how the presence of the LL-TCK leader causes a considerable increase in the expression level of the gene *uid*A (up to 15 times) compared with the original construct.

The analysis of variance allowed to establish that the differences found between the two populations of tobacco considered (transformed with pSTART and pBI121) are statistically significant, as are the differences between the best expressors of the two groups.

In order to exclude effects deriving from epigenetic variations, the analysis was repeated on the T_2 progenies obtained from the self-fertilization of the best primary transformants. In this case too, the plants transformed with pSTART showed expression levels for the gene *uid*A that were much higher than those obtained with pBI121. In particular, an increase equal to 8.6 times was seen in the activity, considering all the plants in their entirety, and equal to 12.5 times considering only the above-average expressors.

In order to determine the effect of LL-TCK on the transcription of the gene *uid*A, T_2 plants were selected (10 plants for pBI121 and 13 plants for pSTART) characterized by intermediate GUS levels for the analysis of the transcript levels by means of real-time RT-PCR. Starting from the total RNA extracted from each plant, the cDNA used as the template in real-time RT-PCR was synthesized. Two pairs of primers were used (one specific for the gene *uid*A and one for the endogenous gene of the 18S RNA) and SYBR-Green PCR Master Mix (Applied Biosystems). The correct quantification was made possible by making 2 calibration lines (one for the transgene and one for the endogenous gene) by means of serial dilutions of control plasmids. The transcription level of the gene *uid*A was then calculated in relative terms for each sample, by means of the percentage ratio between the quantity of mRNA detected for the transgene and the corresponding quantity of ribosomal 18S RNA.

This analysis allowed to verify in the pSTART plants an average transcript level for the gene *uid*A 1.7 times higher than that found in the pBI121 plants.

For 7 pairs of pSTART and pBI121 plants characterized by similar transcript values, the TEI (translational efficiency index) was calculated. The TEI is equivalent to the ratio between the GUS protein value measured with the fluorimetric assay and the relative standardized mRNA value determined by the real-time RT-PCR. By comparing the TEI, it is clear that the new LL-TCK leader not only has an effect on the mRNA levels, but also causes an increase in the translation efficiency of the mRNA.

The LL-TCK sequence allows to increase the expression level of a heterologous protein by acting both on the level of the mRNA content relating to the gene concerned, and also on the level of the final quantity of protein present.

The LL-TCK effect was studied in tobacco, using the constitutive CaMV 35S promoter and the gene *uid*A encoding for the enzyme beta-glucuronidase (GUS), but other uses are possible, in combination with other promoters and other genes.

Although in the examples herein reported the LL-TCK leader is used in combination with the CaMV 35S promoter for enhancement of *uid*A expression in tobacco plants, said leader was successfully used also in tobacco and potato downstream the light-inducible rbcS1 promoter (GenBank Acc. No. AY163904) and in rice downstream the endosperm-specific, phase-dependent glub4 promoter (GenBank Acc. No. AY427571). The genes used in these experiments were those encoding the murine BCL1 antibody, the human beta-glucosidase and a synthetic elastin-like polymer. Since no loss of functionality was recorded in experiments carried out with unrelated genes characterized by a different length, base composition and structure put under the control of promoters with a disparate transcriptional activity and expressed in dicot as well as monocot species, it can be stated that the utility of the LL-TCK leader or similarly composed 5'-UTRs is general, that is, not limited to combinations with certain promoters and/or coding sequences and not limited to certain host species. Therefore, the preferred embodiments of the present invention are comprised in a range of biotechnological applications, including the resistance to biotic/abiotic stresses and herbicides, the production of biofuels, bioplastics, synthetic biopolymers and industrial enzymes, the molecular farming of biopharmaceuticals (e.g. antibodies and their fragments, vaccines, human enzymes, cytokines and growth factors), the improvement of food, feed and fiber quality, the development of reporter and marker gene systems.

Furthermore, it comes within the field of the present invention to construct, inside plant expression vectors, 5'-UTRs in which the following elements are simultaneously present: CaMV 35S *Inr* site, poly(CAA)n, octamer ACAATTAC, poly(CT)n, where n is any number greater than or equal to 2.

All the possible combinations of the elements that constitute the leader sequence 5'-UTR as expressed above, or in the relative variants, irrespective of their relative positioning 5'-3', come within the field of the present invention

Furthermore, the present invention also concerns the sequences complementary to those described above, or the relative variants.

According to a variant, the sequence according to the present invention has a length comprised between 20 and 200 nucleotides, preferably between 40 and 150 nucleotides.

According to a variant, the sequence according to the present invention has a G+C content of less than 60%, preferably less than 50%.

One or more amplification primers also come within the field of the present invention, comprising a nucleotide sequence selected from a group comprising the nucleotide sequences shown in SEQ ID NOS: 2-7 or a complement thereof.

According to another form of execution of the present invention, the sequence according to the present invention can be obtained by:

a) artificial synthesis;

b) natural or induced processes of recombination or mutation inside natural or artificial sequences.

One feature of the present invention also concerns a method for the artificial synthesis of a sequence as described above, using one or more of said amplification primers.

The natural 5'-UTR leader sequences that may be discovered and that appear, to a person skilled in the art, to be non-significant variants, provided they are functionally similar, of the sequence according to the present invention, also come within the field of the present invention.

Sequences deriving from mutation processes of the sequence according to the present invention which appear, to a person of skill, to generate non-significant variants, provided they are functionally similar, of the sequence according to the present invention, are also part of the present invention; the mutations concern irrespectively deletions, insertions, transitions, transversions of one or more nucleotides in the sequence according to the present invention or in the sequence complementary thereto.

The present invention also concerns the bacterial strains carrying plasmids containing the sequence according to the present invention, with particular reference to the species *Escherichia coli, Agrobacterium tumefaciens* and *Agrobacterium rhizogenes.*

The present invention also concerns engineered bacterial strains containing the sequence according to the present invention, irrespective of the type of host organism.

Furthermore, plant cells transformed with expression vectors containing the sequence according to the present invention under the control of a constitutive promoter also come within the field of the present invention.

The following are also covered by the present invention:

- plant cells transformed with expression vectors containing the sequence according to the present invention under the control of a tissue-specific promoter and in particular seed-specific;
- plant cells transformed with expression vectors containing the sequence according to the present invention under the control of an inducible promoter;
- plant cells transformed with expression vectors containing the sequence according to the present invention under the control of a promoter with phase-dependent transcriptional activity;
- plant cells transformed with expression vectors containing the sequence according to the present invention under the control of a promoter active in the chloroplast;
- plant cells transformed with expression vectors containing the sequence according to the present invention under the control of a promoter active in the mitochondrion.

The present invention also comprises plants characterized by the transient expression of any protein whose messenger RNA contains the sequence according to the present invention, transient expression being taken to mean the production of said protein by means of viral vectors, agroinfiltration, electroporation, particle delivery.

The present invention also concerns dicotyledonous plants, with particular reference, but not exclusively, to the species belonging to the families of Solanaceae, Papilonaceae and Cruciferae, stably transformed with expression vectors containing the sequence according to the present invention, and also the progenies of said dicotyledonous plants.

The present invention also concerns monocotyledonous plants, with particular reference, but not exclusively, to the species belonging to the family of

Graminaceae (Poaceae), transformed with expression vectors containing the sequence according to the present invention, and also the progenies of said monocotyledonous plants.

The present invention has an advantageous industrial application, since it also concerns the use of the sequence according to the present invention for one or another of the following activities:

- the biotechnological production of molecules;
- the synthesis of recombinant proteins;
- the synthesis of recombinant proteins intended to induce resistance to viral, bacterial or fungal pathogens;
- the synthesis of recombinant proteins intended to induce resistance to herbicides;
- the synthesis of recombinant proteins intended to obtain an altered composition in fatty acids in the raw material and products deriving therefrom;
- the synthesis of recombinant proteins intended to obtain an altered nutritional value of the raw material and products deriving therefrom;
- the synthesis of recombinant proteins intended for the production of fuels, rubbers and bioplastics;
- the synthesis of industrial enzymes and commercial proteins;
- the synthesis of pharmaceutical proteins;
- the synthesis of orally administered vaccines, intended for men and animals;
- the synthesis of injectable vaccines, intended for men or animals;
- the synthesis of patient-specific injectable vaccines, preferably idiotype-specific, to be used in treating tumors of the lymphatic system;
- the synthesis of proteins involved in the production of secondary metabolites;
- the synthesis of proteins used directly or indirectly as factors to identify and/or select transformed cells.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other characteristics of the present invention will become apparent from the following description of some preferential forms of embodiment, given as a non-restrictive example with reference to the attached drawings wherein:

- fig. 1 is a comparison between the leader sequences in pBI121 and in pSTART, wherein the transcription start site is underlined. Since the sequences between the *Eco* RV site and the transcription start site and between the *Xba* I site and the *uid*A ATG triplet are identical in pSTART and in pBI121, they have been partly omitted (dots);

- fig. 2 shows the expression levels of beta-glucuronidase (GUS) in the transgenic T_1 plants;
- fig. 3 shows the expression levels of beta-glucuronidase (GUS) in the transgenic T_2 plants; the plants are grouped into four groups, each representing sister plants which descend from the best T_1 transformants. Least Significant Difference (P=0.01) = 4.7 U/mg of total protein;
- fig. 4a shows the relative transcript levels of *uid*A (gusA) as determined by real-time RT-PCR in T_2 plants obtained with pSTART and pBI121 characterized by intermediate beta-glucuronidase expression levels. The seven pairs of plants with similar transcript levels are identified;
- fig. 4b shows the values of the translation efficiency index (TEI) for the T_2 plants with similar transcript levels. TEI was calculated as follows: for each transformant, the concentration of beta-glucuronidase (GUS) [U/mg of total protein] was divided by the relative standardized level of mRNA; the highest TEI was considered equal to 1.00 and the values recorded for each transgenic plant were expressed accordingly;
- fig. 5 is a diagram of the overlapping of reverse and forward primers for the synthesis of LL-TCK by means of recursive PCR.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

A) Synthesis of the artificial leader sequence LL-TCK shown in SEQ ID NO: 1.

A.1) The synthesis of the LL-TCK sequence or similarly composed 5'-UTRs can most conveniently be achieved by artificial synthesis, making use of specialized services available on the market. Due to the limited length of the sequence, it is especially useful to add at each side of the leader a flanking region ending with a restriction site already present within the promoter sequence (5' flanking region) and the coding sequence (3' flanking region). It is obvious to a person skilled in the art that these flanking regions will precisely reproduce the sequence upstream the initiator (*Inr*) site and the coding sequence, respectively, unless a modification of the promoter and/or the coding sequence is concurrently planned.

A.2) Another procedure to obtain said leader sequence is recursive PCR

(Podromou and Pearl 1992 Protein Eng., 5, 827-829, Wheeler et al. 1996 Gene, 169, 251-255, Prytulla et al. 1996 FEBS Letters, 399, 283-289).

Once the LL-TCK sequence or a similarly composed 5'-UTR is obtained by either method, leader variants can easily be produced by PCR or any other procedure developed for random or in situ mutagenesis.

In this example, the description of the LL-TCK synthesis by recursive PCR for its insertion into pBI121 (GenBank accession no. AF485783), in particular between the CaMV 35S promoter and the *uid*A coding sequence, is reported.

Five synthetic oligonucleotides were used as primers, having a length comprised between 42 and 54 nt and a partial overlapping degree equal to 24 nt, and a terminal reverse primer of 19 nt, shown respectively in the sequences SEQ ID NOS: 2, 3, 4, 5, 6 and 7.

All primers are written in the 5'-3' direction. The sequences SEQ ID NOS: 2, 4, 6 are forward primers, while the sequences SEQ ID NOS: 3, 5 and 7 are reverse primers. Forward and reverse primers overlap each other according to the diagram in fig. 5.

To facilitate the handling and the insertion of LL-TCK into the vector sequence concerned, a terminal portion starting with an *Eco* RV site was added in 5', while a single *Xba* I site was added to the 3' edge.

Therefore, the primers were designed so as to provide the reconstruction of the portion 3'-terminal of the promoter 35S (from the *Eco* RV site to the *Inr* region) in order to facilitate the subsequent insertion into the vector pBI121 (Clontech).

The external reverse primer introduces the *Xba* I site to the terminal 3', whereas in 5' the *Eco* RV site is used.

In pBI121, these sites fall inside the CaMV 35S promoter and in proximity with the translation start signal of *uid*A, respectively. Therefore, the synthesis was provided of the desired sequence and the cloning to replace the fragment [*Eco* RV - *Xba* I].

The primers comprising the nucleotide sequences shown in SEQ ID NOS: 2-7 were designed and created to confirm the promoter sequence in the region between the *Eco* RV site and the initiator site of CaMV 35S, to synthesize the LL-TCK leader and provide a molecular hook to the terminal 3' for cloning.

The synthesis of LL-TCK was performed by a single PCR, using a PCR reaction mixture in which the concentration of the external primers, SEQ ID NOS: 2 and 7 (corresponding to the two ends of the synthesized segment) was 100 times greater than that of the internal primers SEQ ID NOS: 3, 4, 5 and 6.

In order to achieve a higher fidelity in DNA synthesis, a proof-reading DNA polymerase was used in combination with a 50% reduction of dNTPs concentration.

The PCR reaction mixture is as follows:

10X Pfu Buffer containing 15 mM Mg^{2+}: 10 microL
Primer SEQ ID NO: 2 [10 microM]: 2 microL
Primer SEQ ID NO: 3 [0.1 microM]: 2 microL
Primer SEQ ID NO: 4 [0.1 microM]: 2 microL
Primer SEQ ID NO: 5 [0.1 microM]: 2 microL
Primer SEQ ID NO: 6 [0.1 microM]: 2 microL
Primer SEQ ID NO: 7 [10 microM]: 2 microL
Pfu DNA polymerase [3 U/microL]: 0.8 microL
dNTPs [2.5 milliM each]: 4 microL
Water to a final volume of 100 microL.

In particular, for DNA synthesis and amplification, the Taq polymerase Pfu (Promega) was used, and the following cycle: 1x(95 °C for 5 min); 40x(95 °C for 15 sec; 48 °C for 30 sec; 72 °C for 20 sec); 1x(72 °C for 7 min).

The PCR product was purified by ethanol precipitation, electrophoresed in a 1% agarose gel in TAE buffer, recovered from gel with the aid of a commercial kit, A-tailed with AmpliTaq Gold™, and ligated into pGEM®-T (Promega) for sequencing on both strands.

The ligation mixture was used to transform competent cells of *Escherichia coli*, strain JM101. The absence of any mismatch between the cloned and the designed sequence was verified by sequencing on double strand.

B) Construction of a plant expression vector harbouring the LL-TCK sequence.

The possible addition of flanking regions or molecular hooks to the LL-TCK sequence or similarly composed 5'-UTRs offers a broad range of cloning solutions in expression vectors of different kind. In this example, the method used to clone the [*Eco* RV - *Xba* I] fragment of Example 1 in substitution of the [Eco RV-Xba I] fragment of pBI121 (GenBank accession no. AF485783) is described. Since pBI121 has multiple *Eco* RV sites beyond that inside the CaMV 35S promoter, the latter promoter was excised from pBI121 (Clontech) (Jefferson et al. 1987) making use of the *Hind* III and *Xba* I restriction enzymes.

The fragment was recovered from 1% agarose gel in TAE buffer and subcloned in pUC18 (Pharmacia, GenBank accession no. L08752), previously digested with the same enzymes.

As we said, this passage was necessary because pBI121 has multiple *Eco* RV sites. The pUC18/35S vector obtained was used to make the new combination of 35S promoter – LL-TCK leader.

The LL-TCK sequence was excised from the pGEM-T vector by means of digestion with *Eco* RV and *Xba* I (NEB), separated from the vector sequence by agarose gel electrophoresis and subsequently recovered from gel with the aid of a commercial kit. The pUC18/35S vector was in turn digested with the same enzymes, treated with alcaline phosphatase (Pharmacia), electrophoresed and recovered from gel as above. Then a ligation reaction was carried out at 4°C for 16 hours in the presence of T4 DNA ligase (Promega). In particular, 3.5 ng of the [*Eco* RV - *Xba* I] fragment were combined with 25 ng of the vector in the presence of 1 U T4 DNA ligase in a volume of 10 microL containing a suitable reaction buffer.

The pBI121 vector (Clontech) was subjected to digestion with *Xba* I and *Hind* III (NEB) to remove the CaMV 35S promoter. The complex 35S-LL-TCK was in turn excised from the cloning vector pUC18 by means of digestion with the same enzymes; the pBI121 vector frame and the 35S-LL-TCK insert were electrophoresed and recovered from gel as above. Finally, a ligation of 35S-LL-TCK in the pBI121 framework was performed, obtaining the vector pBI121/35S-LL-TCK:: *uid*A::NOS to which the name pSTART (fig. 1) was assigned.

C) Transformation of plants with an expression vector containing the LL-TCK sequence.

Transgenic plants harbouring the LL-TCK sequence or similarly composed 5'-UTRs can be produced through a range of methods, including coinfection with engineered strains of *Agrobacterium* spp., infection or trasfection with engineered strains of phytoviruses, electroporation, particle delivery, DNA microinjection.

The pSTART expression vector was electroporated into *Agrobacterium tumefaciens* strain EHA 105 and the transformed *Agrobacterium* cells used for tobacco (*Nicotiana tabacum* L., cv. Xanthi) transformation. Briefly, 2 ml LB medium supplemented with kanamycin (50 mg/L) were inoculated with transformed *Agrobacterium* cells. Bacterial cultures were incubated at 29 °C for 16 hours. Leaf discs (7 mm in diameter) were obtained with a cork borer from axenically grown, 30-d old seedlings or from mature leaves collected from plants at the late-rosette stage. In the latter case, tobacco leaves were rinsed with distilled water, surface-sterilized in 1% sodium hypochlorite for 5 min and in 95% ethanol for 30 sec and blotted to sterile filter paper under a laminar flow hood.

Leaf discs were placed in a Petri dish containing 15 mL of Murashige and Skoog medium supplemented with 0.1 mg/L naphthalene acetic acid (NAA), 1 mg/L 6-benzyladenine (BA), 30 g/L sucrose and solidified with 8 g/L agar. Immediately after this transfer, 2 ml of the above-mentioned *Agrobacterium* culture were poured in the Petri dish and the leaf discs were uniformly wetted. After removal of the LB medium in excess, the dish was sealed and incubated for 24 hours at 25 °C in the light (30.5 microE/square meter/sec).

Leaf discs were then transferred to a new Petri dish containing 15 mL of Murashige and Skoog medium supplemented with 0.1 mg/L naphthalene acetic acid (NAA), 1 mg/L 6-benzyladenine (BA), 500 mg/L cefotaxime, 30 g/L sucrose and solidified with 8 g/L agar. They were incubated for a week at 28 °C and illuminated for 16 hours/day; they were eventually transferred to a substrate identical to the former except for the presence of 200 mg/L kanamycin. Explants were subcultured every 3 weeks; regenerated shoots were isolated from callus tissue and rooted on semisolid Murashige and Skoog medium supplemented with 2 mg/L indole-3-butyric acid, 500 mg/L cefotaxime, 200 mg/L kanamycin, 30 g/L sucrose.

Putative transgenic plants were potted in peat and grown in a greenhouse under Powerstar HQI-T lamps (Osram) (200 mM photons/square meter/sec at canopy level) for 16 hours/day at 25 °C/19 °C light/dark.

In this example, transformation was confirmed by PCR and beta-glucuronidase assays. For PCR assay, total DNA was extracted according to

Doyle and Doyle (1990) and the following primers were used:

Forward 5' - ACAATTACGTATTTCTCTCTCTAGA - 3'

Reverse 5' - CGATCGGGGAAATTCGAGCTC - 3'

The forward and reverse primers anneal to the end of the LL-TCK sequence and to part of the NOS terminator, respectively and do not give rise to any amplification product in untransformed Xanthi plants (in transgenic plants the amplicon length is 1936 bp, as expected).

About 93% of the regenerated plants were found transgenic when a standard reaction mix was formed and the following temperature cycling was used: 1x(94 °C for 5 min); 40x(94 °C for 1 min 15 sec; 60 °C for 45 sec; 72 °C for 2 min); 1x(72 °C for 5 min).

Plant transformation was further demonstrated by GUS histochemical assays (Jefferson et al. 1987) and fluorimetric determination of GUS activity. Controls consisted of Xanthi plants raised in vitro from uninfected discs. The methods used for the fluorimetric assay and the results obtained are reported in detail in point D.

The same procedure was followed to produce and characterize transgenic plants harbouring the original leader sequence. *Ceteris paribus*, no effect of the LL-TCK sequence was noted upon regeneration and transformation rates.

D) Effect of the LL-TCK sequence on transgene expression levels.

As previously indicated, the plasmids pBI121 and pSTART were individually used for *Agrobacterium*-mediated transformation of tobacco leaf discs. Since in both cases the gene under control is *uid*A (also known as *gus*A), transgene expression levels achievable with LL-TCK and the widely-distributed pBI121 leader can be compared directly by determining the activity of the *uid*A encoded enzyme, beta-glucuronidase (EC 3.2.1.31). About twenty primary transformants (that is transgenic plants belonging to the first generation, T_1) of each population were assayed for transgene presence by PCR and subsequently analysed for beta-glucuronidase activity (fig. 2). When the late rosette stage was reached (30 days upon hardening), the 3 youngest leaves were collected from each plant to obtain crude sap by pressing (Erich Pollähne); 100 microL of crude sap were mixed with 2 volumes of extraction buffer (Jefferson 1989) containing 12 mg of high molecular weight polyvinyl pyrrolidone (PVP). After centrifugation for 15 min at

11,500xg, the supernatant was collected and analysed fluorimetrically (Dyna Quant 200 fluorometer; GE Healthcare) in duplicate using 4-methylumbelliferyl-beta-D-glucuronide (MUG; Sigma-Aldrich) as substrate. Trials were made to determine the background noise due to intrinsic fluorescence of the samples, quenching, as well as substrate degradation by factors other than the recombinant enzyme. The level of transgene expression was measured in terms of beta-glucuronidase units per mL of crude sap, one unit being defined as the amount of enzyme releasing 1 nM 4-methylumbelliferone (4-MU) min^{-1} using the same assay conditions as previously described (Jefferson 1989). Data were submitted to a log transformation to avoid any correlation between variance and mean; the analysis of variance was carried out after checking the normality of the distribution of log data by means of the Kolmogorov-Smirnov test and the homogeneity of variances with Bartlett's formula. Averages were compared with Duncan's multiple range test at the probability level, P=0.05.

The analysis of variance carried out on fluorescence data showed the absence of any significant variation among young leaves of the same plant, whereas remarkable differences existed among plants. Specifically, the synthetic leader determined a highly significant increase (up to 15-fold) in *uid*A expression (Table 1).

Table 1: Beta-glucuronidase activity (U/mL of crude sap) in randomly-chosen T_1 plants

pSTART	pBI121
96.10	6.24
89.81	5.37
67.06	5.34
57.89	4.52
33.16	4.20
27.62	3.55
15.97	2.95
4.24	2.89
3.51	2.79
1.85	2.34

1.64	2.30
1.46	1.62
1.21	1.17
0.77	1.12
0.64	0.75
0.46	0.47
0.38	0.23
	0.15

To demonstrate that these outcomes were not biased by epigenetic variation, analyses were repeated on T_2 progenies. In particular, the best 4 T_1s in each population were selfed and the resulting seed plated on a kanamycin-enriched medium for selection; 5-7 resistant T_2 plants were randomly chosen within each progeny and raised to the late rosette stage for verification of the transgene presence by PCR and measurement of beta-glucuronidase activity by fluorimetric assays; enzyme levels were again expressed as units per mL of crude sap but also as units per mg total protein (as determined by Bradford assay) to account for plant-to-plant variation in metabolism.

Similarly to what observed in the T_1 generation, transgenic T_2 plants harbouring the new leader showed significantly higher levels of *uid*A expression (fig. 3); as compared to the pBI121 leader, a 8.6-fold and a 12.5-fold increase of activity were estimated taking into account the whole plant population or the above-average expressors, respectively (Table 2).

Table 2: Beta-glucuronidase activity (U/mg protein) in T_2 plants obtained from the best 4 T_1s

pSTART	pBI121
139.07	789
71.27	5.17
45.43	4.44
44.81	3.96
41.67	3.94
32.92	3.78

29.81	3.77
19.79	3.58
18.04	3.55
16.40	3.23
15.10	3.15
13.55	3.09
12.84	2.94
12.32	2.62
10.78	2.45
10.02	1.71
9.00	1.28
8.50	1.22
6.98	1.14
6.57	0.64
5.39	0.52
5.16	0.46
5.01	0.44
3.17	0.31

E) Effect of the LL-TCK sequence on gene transcription and translation.

The combination of different elements in LL-TCK or similarly composed 5'-UTRs reflects in measurable improvements of transcriptional as well as translational efficiency of a given transgene. In this example, such improvements are shown in transgenic T_2 tobacco plants obtained with pBI121 and pSTART, as described before. In order to collect such evidence, plants belonging to the pBI121 or the pSTART group (10 and 13 T_2s, respectively) and characterized by intermediate *uid*A expression levels were analysed to determine:

i. the mean transcript levels of *uid*A;
ii. the mean transcript levels of 18S RNA;
iii. the amount of beta-glucuronidase actually produced.

To minimize experimental error, one young leaf was collected from each plant and cut in two, one half being used for RNA isolation, the other for beta-glucuronidase assay.

Total RNA was extracted with RNAgents Total RNA Isolation System (Promega). First-strand cDNA was synthesized from 1 microg RNA by AMV Reverse Transcriptase (Promega) in the presence of random primers. The cDNA synthesis reaction was diluted 1:5 and 1 microL was used for real-time PCR (qRT-PCR).

qRT-PCRs were performed making use of SYBR-Green PCR Master Mix (Applied Biosystems) and specific primers each at a 0.3 microM final concentration. All reactions were carried out with the iCycler iQ multicolor real-time PCR detection system (Bio-Rad) and run with the following program: 1x(95 °C for 10 min); 50x(95 °C for 15 sec; 62 °C for 30 sec; 72 °C for 30 sec). To amplify the *uid*A transcript, the following primers were used:

Forward 5'-TTACGCTGAAGAGATGCTCGAC-3'

Reverse 5'-CCTAAAGAGAGGTTAAAGCCGACAG-3'

For the 18S RNA target sequence, the primers were designed on the basis of GenBank accession n. AJ236016:

Forward 5'-ACATCCAAGGAAGGCAGCAG-3'

Reverse 5'-GACTCATAGAGCCCGGTATTGTTATT-3'

In both cases, the amplicon length was 90 bp. In each PCR run, serial dilutions of control plasmids were included in parallel with known amounts of input copy number in order to draw standard calibration curves. Specifically, 10-fold serial dilutions (from 10^5 to 10^2 copies) of the *uid*A -harbouring plasmid pBI221, were used as templates. For the same purpose, a 550bp fragment of the 18S RNA gene (AJ236016) was cloned in pGEM-T Easy (Promega) and used in the range of 10^8 –10^5 copies.

The starting quantities of *uid*A transcript and control RNA were determined with ICycler IQ real-time detection system software ver. 3.0. For each sample, at least 3 independent estimates were performed; in all cases, the maximum value for the variation coefficient (VC=SD/mean) was fixed at 20%. The % ratio between the mean transcript levels of *uid*A and 18S RNA was calculated for each sample. To standardize data, the highest *uid*A transcript level was considered equal to 1.00 and values recorded for each transgenic plant were expressed accordingly.

Results obtained in qRT-PCR indicated that the substitution of the pBI121 leader with LL-TCK determines a clear increase of average *uid*A transcript levels. Specifically, the transcriptional efficiency was found 1.7-fold higher in plants harbouring the LL-TCK leader (Table 3).

Table 3: % ratio between *uid*A and 18S transcripts in several T_2 plants

pSTART	pBI121
0.01371%	0.00881%
0.01189%	0.00832%
0.01132%	0,00761%
0.00968%	0,00437%
0.00960%	0,00315%
0.00896%	0,00308%
0.00578%	0,00282%
0.00556%	0,00234%
0.00425%	0.00225%
0.00315%	0.00158%
0.00272%	0.00072%
0.00201%	
0.00189%	

In addition, after having identified 7 plant couples with nearly overlapping *uid*A transcript levels (fig. 4a), a translational efficiency index (TEI) was attributed to each plant (fig. 4b), making the ratio between beta-glucuronidase concentration and the relative standardized level of *uid*A transcript. By comparison of TEI values, the two leaders were found to determine a clearly different translational efficiency of *uid*A transcripts, which was greater using the leader according to the present invention (Table 4).

Table 4: ratio between GUS enzymatic activity and relative transcription level of *uid*A gene measured in some T_2 plants

TEI	pSTART	pBI121	TEI
1.00	299.07	103.06	0.34
0.93	277.07	66.58	0.22

0.71	211.87	58.52	0.20
0.68	203.76	19.82	0.07
0.65	195.24	42.97	0.14
0.65	194.12	47.18	0.16
0.52	154.91	20.86	0.07
0.34	102.01	29.52	0.10
0.30	89.48	8.94	0.03
0.27	80.26	12.34	0.04
0.25	73.69	22.81	0.08
0.20	59.70		
0.17	49.40		

SEQUENCE LISTING

<110> UNIVERSITA' DEGLI STUDI DI UDINE

<120> ARTIFICIAL DNA SEQUENCE WITH OPTIMIZED LEADER FUNCTION IN 5' (5'-UTR) FOR THE IMPROVED EXPRESSION OF HETEROLOGOUS PROTEINS IN PLANTS

<130> T3-2288

<150> IT UD2006A0000280
<151> 2006-12-29

<160> 7

<170> PatentIn version 3.3

<210> 1
<211> 73
<212> DNA
<213> Artificial

<220> 5'UTR
<223> improved expression of heterologous proteins in plants

<400> 1
acacgtattt ttacaacaat accaacaaca acaacaacaa acaacattac aattacgtat 60

ttctctctct aga 73

<210> 2
<211> 54

<212> DNA
<213> Artificial

<220> primer_bind
<223> forward primer element for artificial synthesis by means of recursive PCR

<400> 2
gatatctcca ctgacgtaag ggatgacgca caatcccact atccttcgca agac 54

<210> 3
<211> 54
<212> DNA
<213> Artificial

<220> primer_bind
<223> reverse primer element for artificial synthesis by means of recursive PCR

<400> 3
atgaaatgaa cttccttata tagaggaagg gtcttgcgaa ggatagtggg attg 54

<210> 4
<211> 53
<212> DNA
<213> Artificial

<220> primer_bind
<223> forward primer element for artificial synthesis by means of recursive PCR

<400> 4
tctatataag gaagttcatt tcatttggag agaacacgta tttttacaac aat 53

<210> 5
<211> 52
<212> DNA
<213> Artificial

<220> primer_bind
<223> reverse primer element for artificial synthesis by means of recursive PCR

<400> 5
taatgttgtt tgttgttgtt gttgttggta ttgttgtaaa aatacgtgtt ct 52

<210> 6
<211> 53
<212> DNA
<213> Artificial

<220> primer_bind
<223> forward primer element for artificial synthesis by means of recursive PCR

<400> 6
aacaacaaca acaaacaaca ttacaattac gtatttctct ctctagagga tcc 53

<210> 7
<211> 19
<212> DNA

<213> Artificial

<220> primer_bind

<223> reverse primer element for artificial synthesis by means of recursive PCR

<400> 7

cacccgggga tcctctaga 19

CLAIMS

1. Nucleotide leader sequence 5'-UTR, characterized in that it comprises elements favorable to gene expression, such as repeated CAA trinucleotide elements in combination with repeated CT dinucleotide elements.

2. Sequence as in claim 1, containing a poly(CAA) region, meaning an oligonucleotide consisting of two or more copies of the CAA element, preferably but not necessarily contiguous with each other.

3. Sequence as in claim 1, containing a poly(CT) region, meaning an oligonucleotide consisting of two or more copies of the CT element, preferably but not necessarily contiguous with each other.

4. Sequence as in claim 1, containing one or more copies of the octamer ACAATTAC.

5. Sequence as in claim 1, comprising a combination of the sequences in claims 2 and 3.

6. Sequence as in claim 1, comprising a combination of the sequences in claims 2 and 4.

7. Sequence as in claim 1, comprising a combination of the sequences in claims 3 and 4.

8. Sequence as in claim 1, comprising a combination of the sequences in claims 2, 3 and 4.

9. Sequence as in claim 1, comprising a combination of the sequences in any claim from 2 to 8 with the *Inr* site of CaMV 35S, that is, the transcription initiation site of the cauliflower mosaic virus 35S promoter.

10. Sequence as in claim 1, comprising the sequence shown in SEQ ID NO: 1.

11. The construction inside plant expression vectors of 5'-UTRs in which the following elements are simultaneously present: *Inr* site of CaMV 35S, poly(CAA)n, octamer ACAATTAC, poly(CT)n, where n is any number greater than or equal to 2.

12. Sequence as in claim 1, comprising all the possible combinations of the elements that constitute the leader sequence 5'-UTR as in any claim hereinbefore, irrespective of their relative positioning 5'-3'.

13. Sequence as in claim 1, comprising sequences complementary to those reported in any one of the previous claims.

14. Sequence as in claim 1, having a length comprised between 20 and 200 nucleotides, preferably between 40 and 150 nucleotides.

15. Sequence as in claim 1, having a G+C content of less than 60%, preferably less than 50%.

16. Sequence as in claim 1 inasmuch as it is obtainable from: a) artificial synthesis; b) natural or induced processes of recombination inside natural or artificial sequences.

17. Amplification primer comprising a nucleotide sequence selected from a group comprising the nucleotide sequences shown in SEQ ID NOS: 2–7 or a complement thereof.

18. Natural leader sequences 5'-UTR possibly discovered which appear non-significant variants, provided they are functionally similar, of the sequence as in any claim hereinbefore.

19. Sequences deriving from mutations of the sequence as in any claim hereinbefore, which appear to generate non-significant variants, provided they are functionally similar, of the sequence as in any claim hereinbefore, wherein the mutations concern irrespectively deletions, insertions, transitions, transversions of one or more nucleotides in the sequence as in any claim hereinbefore or in the sequence complementary thereto.

20. Method for the artificial synthesis of a sequence as in any claim from 1 to 19, characterized in that it provides to use one or more primers as in claim 17.

21. Bacterial strains carrying plasmids containing the sequence as in any claim hereinbefore, with particular reference to the species *Escherichia coli, Agrobacterium tumefaciens* and *Agrobacterium rhizogenes.*

22. Engineered virus strains containing the sequence as in any claim from 1 to 19, irrespective of the host organism.

23. Plant cells transformed with expression vectors containing the sequence as in any claim from 1 to 19 under the control of a constitutive promoter.

24. Plant cells transformed with expression vectors containing the sequence as in any claim from 1 to 19 under the control of a tissue-specific promoter and in particular seed-specific.

25. Plant cells transformed with expression vectors containing the sequence as in any claim from 1 to 19 under the control of an inducible promoter.

26. Plant cells transformed with expression vectors containing the sequence as in any claim from 1 to 19 under the control of a promoter with phase-dependent transcriptional activity.

27. Plant cells transformed with expression vectors containing the sequence as in any claim from 1 to 19 under the control of a promoter active in the chloroplast.

28. Plant cells transformed with expression vectors containing the sequence as in any claim from 1 to 19 under the control of a promoter active in the mitochondrion.

29. Plants characterized by the transient expression of any protein in which the RNA messenger contains the sequence as in any claim from 1 to 19, transient expression being taken to mean the production of said protein by means of viral vectors, agroinfiltration, particle delivery, electroporation.

30. Dicotyledonous plants stably transformed with expression vectors containing the sequence as in any claim from 1 to 19.

31. Plants as in claim 30 comprising one or more species belonging to the families Solanaceae, Papilonaceae and Cruciferae.

32. Progenies of the dicotyledonous plants as in claim 30 or 31.

33. Monocotyledonous plants transformed with expression vectors containing the sequence as in any claim from 1 to 19.

34. Plants as in claim 33 comprising one or more species belonging to the family Graminaceae (Poaceae).

35. Progenies of the monocotyledonous plants as in claim 33 or 34.

36. Use of the sequence as in any claim from 1 to 19 for the biotechnological production of molecules.

37. Use of the sequence as in any claim from 1 to 19 for the synthesis of recombinant proteins.

38. Use of the sequence as in any claim from 1 to 19 for the synthesis of recombinant proteins intended to induce resistance to viral, bacterial or fungal pathogens.

39. Use of the sequence as in any claim from 1 to 19 for the synthesis of recombinant proteins intended to induce resistance to herbicides.

40. Use of the sequence as in any claim from 1 to 19 for the synthesis of recombinant proteins intended to obtain a modified composition in fatty acids in

the raw material and the products deriving therefrom.

41. Use of the sequence as in any claim from 1 to 19 for the synthesis of recombinant proteins intended to obtain a changed nutritional value of the raw material and the products deriving therefrom.

42. Use of the sequence as in any claim from 1 to 19 for the synthesis of recombinant proteins intended for the production of fuels, rubbers and bioplastics.

43. Use of the sequence as in any claim from 1 to 19 for the synthesis of industrial enzymes and commercial proteins.

44. Use of the sequence as in any claim from 1 to 19 for the synthesis of pharmaceutical proteins.

45. Use of the sequence as in any claim from 1 to 19 for the synthesis of orally administered vaccines, intended for man or animals.

46. Use of the sequence as in any claim from 1 to 19 for the synthesis of injectable vaccines, intended for men or animals.

47. Use of the sequence as in any claim from 1 to 19 for the synthesis of patient-specific injectable vaccines, preferably idiotype-specific, to be used in the treatment of tumors of the lymphatic system.

48. Use of the sequence as in any claim from 1 to 19 for the synthesis of proteins involved in the production of secondary metabolites.

49. Use of the sequence as in any claim from 1 to 19 for the synthesis of proteins usable directly or indirectly as factors to identify and/or select transformed cells.

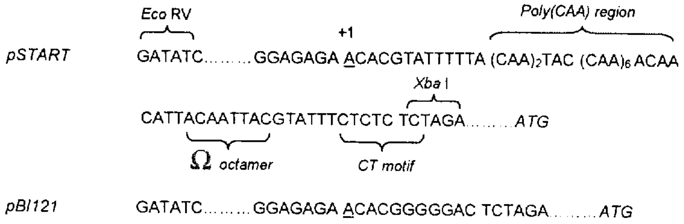


Fig. 1

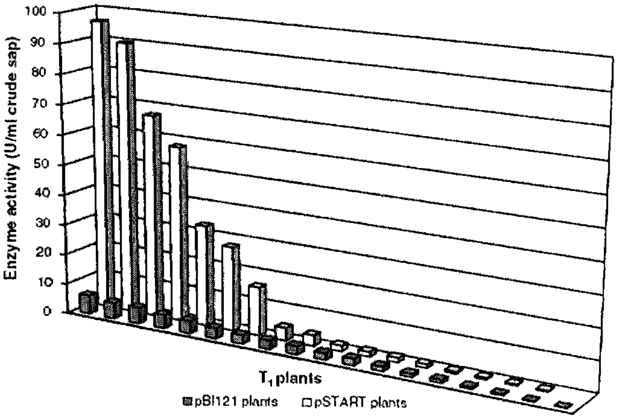


Fig. 2

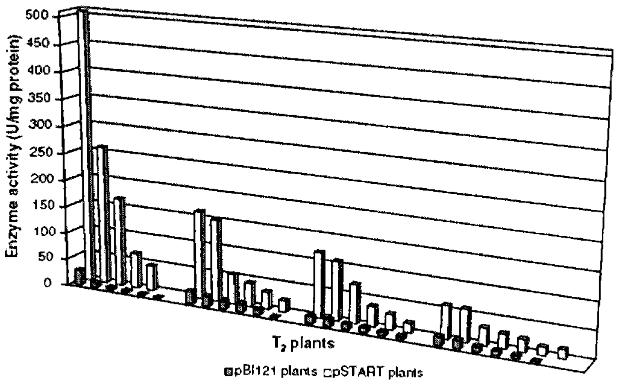


Fig. 3

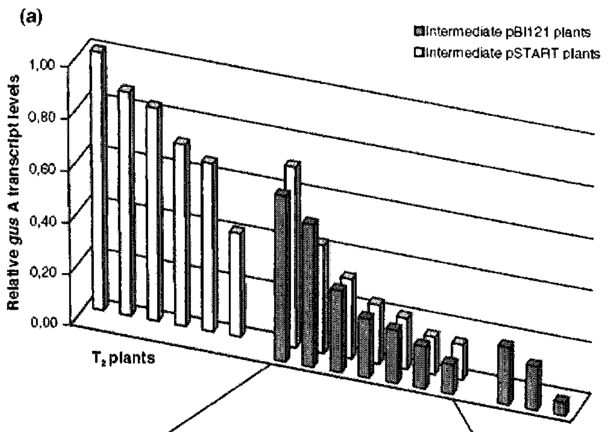


Fig. 4a

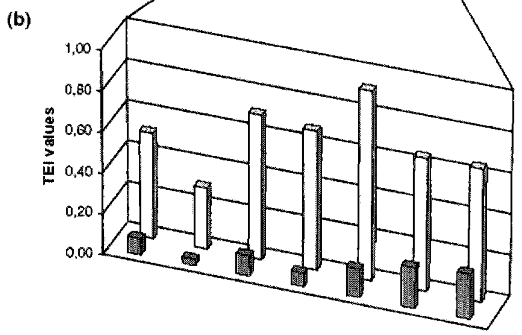


Fig. 4b

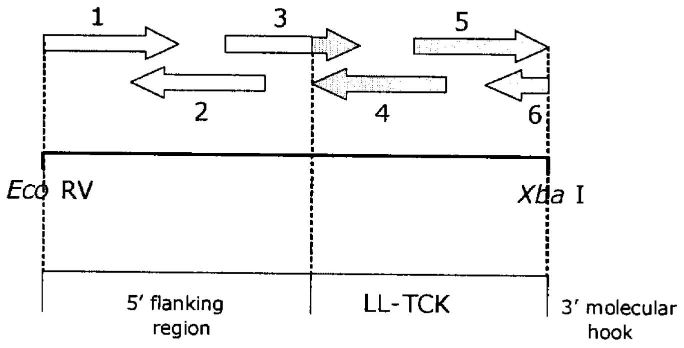


Fig. 5