Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2573986 A1 2006/02/09

(21) 2 573 986

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2005/07/30
(87) Date publication PCT/PCT Publication Date: 2006/02/09
(85) Entrée phase nationale/National Entry: 2007/01/15
(86) N° demande PCT/PCT Application No.: EP 2005/008285
(87) N° publication PCT/PCT Publication No.: 2006/013072
(30) Priorités/Priorities: 2004/08/02 (US60/598,001);
2005/07/01 (US60/696,209)

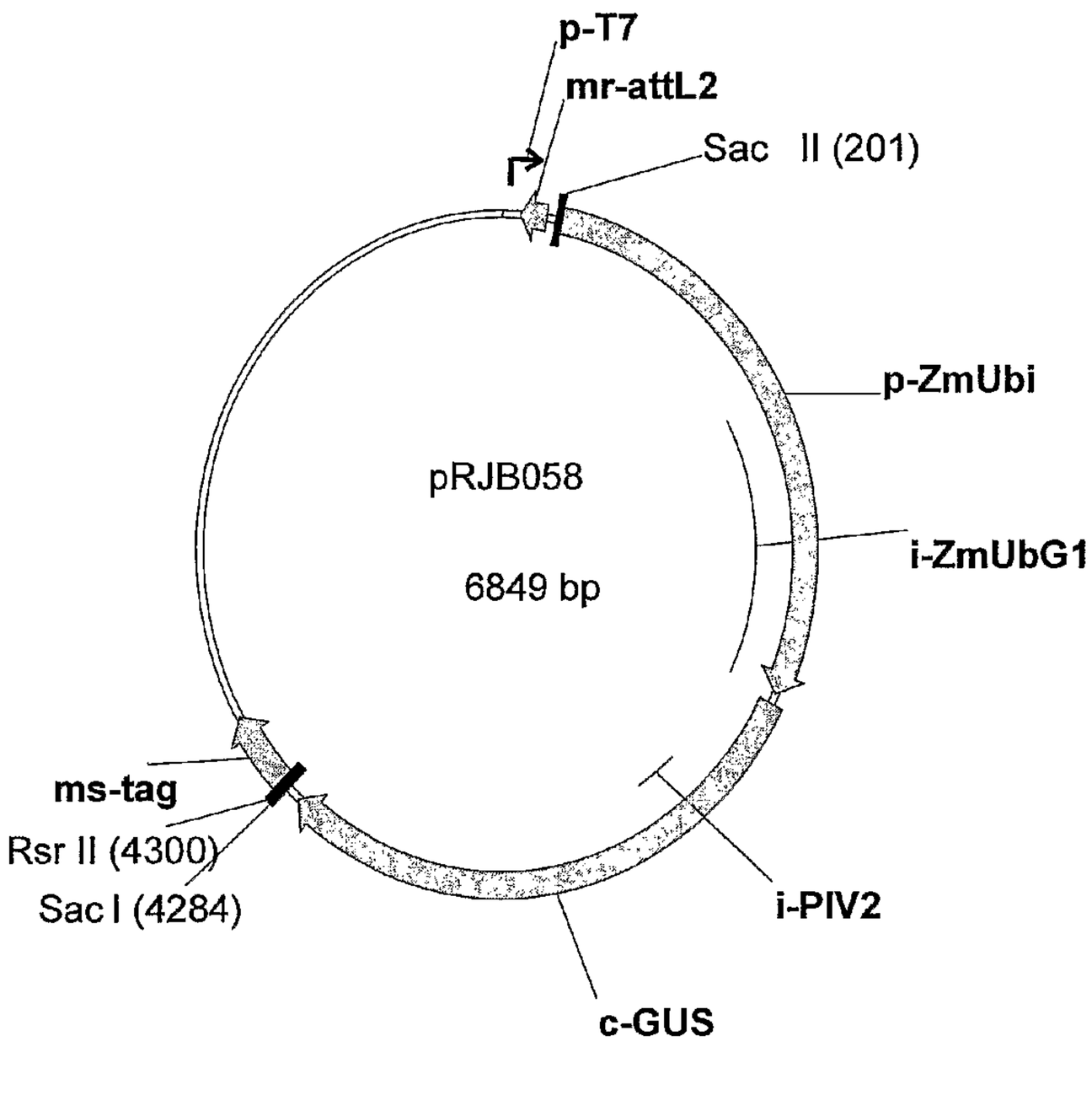
(51) Cl.Int./Int.Cl. *C12N 15/10* (2006.01),
*C12N 15/82* (2006.01)

(71) Demandeur/Applicant:
BASF PLANT SCIENCE GMBH, DE

(72) Inventeurs/Inventors:
SONG, HEE-SOOK, US;
KOCK, MICHAEL, DE;
BROWN, JEFFREY A., US;
LOYALL, LINDA PATRICIA, DE;
XING, LIQUN, US;
JIA, HONGMEI, US;
MCMILLAN, JOHN, US;
IRELAND, LESLEY, US

(74) Agent: ROBIC

(54) Titre : PROCEDE POUR L'ISOLEMENT DE SEQUENCES DE TERMINAISON DE TRANSCRIPTION
(54) Title: METHOD FOR ISOLATION OF TRANSCRIPTION TERMINATION SEQUENCES




(57) Abrégé/Abstract:
The invention relates to efficient, high-throughput methods, systems, and DNA constructs for identification and isolation of transcription termination sequences. The invention relates further to specific terminator sequences identified by said methods isolated from rice.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 February 2006 (09.02.2006)

PCT

(10) International Publication Number
WO 2006/013072 A3

(51) International Patent Classification:
*C12N 15/10* (2006.01) *C12N 15/82* (2006.01)

(21) International Application Number:
PCT/EP2005/008285

(22) International Filing Date: 30 July 2005 (30.07.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/598,001 2 August 2004 (02.08.2004) US
60/696,209 1 July 2005 (01.07.2005) US

(71) Applicant *(for all designated States except US)*: **BASF Plant Science GmbH** [DE/DE]; 67056 Ludwigshafen (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SONG, Hee-Sook** [KR/US]; 1504 Pineview Drive, Raleigh, NC 27606 (US). **KOCK, Michael** [DE/DE]; Am Leutbusch 12, 67105 Schifferstadt (DE). **BROWN, Jeffrey A.** [US/US]; 1005 Modest Way, Apex, NC 27502 (US). **LOYALL, Linda Patricia** [DE/DE]; Fred-Joachim-Schoeps-Str. 17, 68229 Mannheim (DE). **XING, Liqun** [CA/US]; 203 Laurens Way, Chapel Hill, NC 27516 (US). **JIA, Hongmei** [CN/US]; 2001 Henniker Street, Apex, NC 27523 (US). **MCMILLAN, John** [US/US]; 8345 Nantahala Drive, Raleigh, NC 27612 (US). **IRELAND, Lesley** [—/US]; 304 Madison Grove Place, Morrisville, NC 27560 (US).

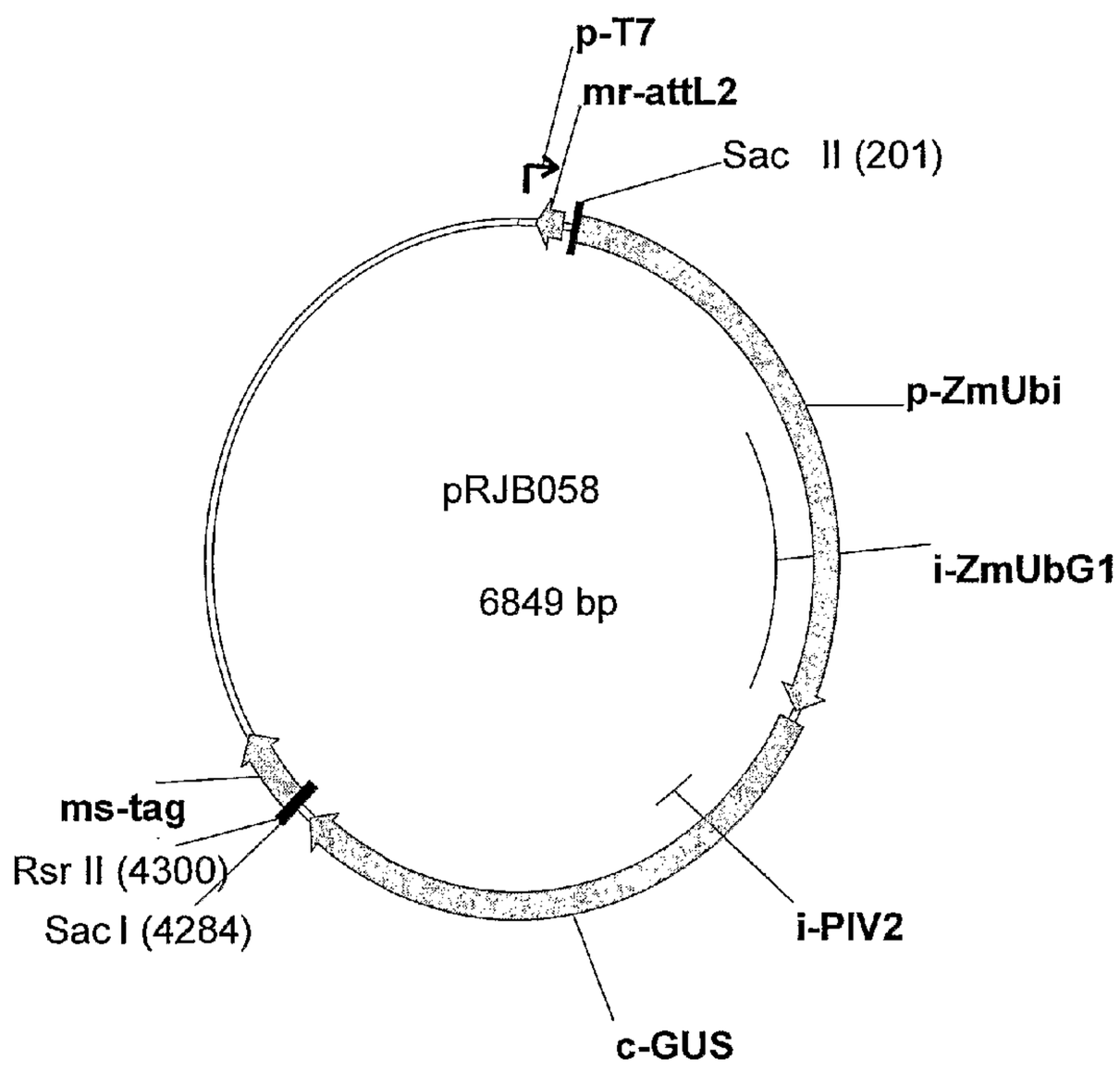
(74) Agent: **PRESSLER, Uwe**; c/o BASF Aktiengesellschaft, 67056 Ludwigshafen (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, NL, PL, PT,

*[Continued on next page]*

(54) Title: METHOD FOR ISOLATION OF TRANSCRIPTION TERMINATION SEQUENCES




(57) Abstract: The invention relates to efficient, high-throughput methods, systems, and DNA constructs for identification and isolation of transcription termination sequences. The invention relates further to specific terminator sequences identified by said methods isolated from rice.

WO 2006/013072 A3

WO 2006/013072 A3



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*

**(88) Date of publication of the international search report:**
3 August 2006

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 109

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

# JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 109

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

# Method for isolation of transcription termination sequences

## FIELD OF THE INVENTION

The invention relates to efficient, high-throughput methods, systems, and DNA constructs for identification and isolation of transcription termination sequences.

## BACKGROUND OF THE INVENTION

The aim of plant biotechnology is the generation of plants with advantageous novel properties, such as pest and disease resistance, resistance to environmental stress (*e.g.*, water-logging, drought, heat, cold, light-intensity, day-length, chemicals, *etc.*), improved qualities (*e.g.*, high yield of fruit, extended shelf-life, uniform fruit shape and color, higher sugar content, higher vitamins C and A content, lower acidity, *etc.*), or for the production of certain chemicals or pharmaceuticals (Dunwell 2000). Furthermore resistance against abiotic stress (drought, salt) and/or biotic stress (insects, fungal, nematode infections) can be increased. Crop yield enhancement and yield stability can be achieved by developing genetically engineered plants with desired phenotypes.

For all fields of biotechnology, beside promoter sequences, transcription terminator sequences are a basic prerequisite for the recombinant expression of specific genes. In animal systems, a machinery of transcription termination has been well defined (Zhao *et al.,* 1999; Proudfoot, 1986; Kim *et al.,* 2003; Yonaha and Proudfoot, 2000; Cramer *et al.,* 2001; Kuerstem and Goodwin, 2003). Effective termination of RNA transcription is required to prevent unwanted transcription of trait-unrelated (downstream) sequences, which may interfere with trait performance (see below for more details). Especially arrangement of multiple gene expression cassettes in local proximity (*e.g.*, within one T-DNA) is often causing suppression of gene expression of one or more genes in said construct in comparison to independent insertions (Padidam and Cao, 2001). This is causing problems especially in cases were strong gene expression from all cassettes is required

Previously efficiency of transcription termination had to be analyzed either by *in vitro* or *in vivo* transcription analysis of individual transcription termination sequences, which is a laborious and time-consuming procedure based on trial-and-error (Yonaha and Proudfoot, 1999, 2000; Yarnell and Roberts, 1999). To simplify this process, single nucleotide-recognizing probe such as beacon has been used for *in vitro* transcription (Liu *et al.,* 2002).

In plants, understanding transcription termination and re-initiation is at the infant stage. There are no clearly defined polyadenylation signal sequences. Hasegawa *et al.* (2003) were not able to identify conserved polyadenylation signal sequences in both *in vitro* and *in vivo* systems in *Nicotiana sylvestris* and to determine the actual length of the primary (non-polyadenylated) transcript. There are vague ideas that weak terminator can generate read-through, which affects the expression of the genes located in neighboring expression cassettes (Padidam and Cao, 2001). Appropriate control of transcription termination will prevent read-through into sequences (*e.g.*, other expression cassettes) localized downstream and will further allow efficient recycling of RNA polymerase II, which will improve gene expression.

Prediction of functional, efficient transcription termination sequences by bioinformatics is not feasible alternative since virtually no conserved sequences exist which would allow for such a prediction. Prediction of the efficiency in transcription termination of such sequences is even more beyond. Furthermore, experimental determination of the actual length and sequence of the primary transcript is difficult since these structures are highly instable being rapidly converted into polyadenylated transcripts (Hasegawa *et al.*, 2003).

Production of genetically modified cells and organisms (such as plants) requires appropriate recombinant DNA in order to introduce genes of interest. The recombinant DNA contains more than one expression cassette, in general. The expression cassette is composed of promoter, gene of interest, and terminator. The expression of the gene of interest in the expression cassette can be negatively affected by inappropriate termination of transcription from the neighboring cassette. Transcriptional read-through and/or multiple use of the same transcription termination sequence may have one or more of the following disadvantages:

1. Unwanted expression of downstream sequences may cause undesirable effects (*e.g.*, changes in metabolic profile, gene silencing *etc.*).
2. Unwanted expression of downstream sequences raises higher hurdles in deregulation proceedings.
3. Multiple use of identical transcription termination sequences may lead to failure of the whole transgenic expression approach by epigenic silencing. Because the present panel of evaluated transcription termination sequences is currently very limited, multiple use of the same transcription termination sequence in one transgenic organism is often unavoidable, which has proofed to result in unintended silencing of the entire transgenic expression constructs (Matzke 1994; Matzke 1989)
4. Enablement of constructs comprising multiple gene expression cassettes without undesired interaction of transcription of different cassettes. Such interactions may – depending on the orientation of the cassettes – include unintended expression (*e.g.*, in case of expression cassettes having the same direction of their reading frames) or unintended gene silencing (*e.g.*, in case of inverted orientation of the cassettes).

In consequence, there is an unsolved demand (especially in the plant biotech area) for tight and alternative transcription termination sequences. There is no easy and reliable screening system to identify "tight" terminators that efficiently terminate transcription. It is therefore an objective of the present invention, to provide a method to easily identify such termination sequences and to provide tight and alternative transcription termination sequences for use on plants. This objective is achieved by this invention.

## BRIEF DESCRIPTION OF THE INVENTION

Accordingly, a first embodiment of the invention related to a method for identification and isolation of transcription termination sequences for comprising the steps of:

i) providing a screening construct or screening vector comprising
   a) a promoter sequence, and
   b) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences, and
   c) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic,

   wherein insertion of an efficient transcription terminator into said insertion site changes expression of said additional sequences by said promoter sequence in comparison to no insertion, and
ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and
iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector, and
iv) introducing said screening construct or screening vector with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and
v) identifying and/or selecting screening construct or screening vector with a changed readily detectable characteristic in comparison to no insertion, and
vi) isolating the inserted DNA sequences from said identified and/or selected screening construct or screening vector for use as transcription termination sequences and – optionally – determining their sequence.

There are various options for localization of said insertion site in relation to said additional sequences. For example the insertion site may preferably be at a position selected from group of:

i) upstream of the additional sequences between said promoter and said additional sequences, and
ii) downstream of the additional sequences, and
iii) in between said additional sequences.

Depending on the localization of the insertion site to said additional sequences several especially preferred embodiments result. In one preferred embodiment method for identification and isolation of transcription termination sequences comprises the steps of:

i) providing a screening construct or screening vector comprising in 5' to 3' direction
   a) a promoter sequence, and
   b) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences, and
   c) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic,

wherein insertion of an efficient transcription terminator into said insertion site suppresses expression of said additional sequences by said promoter sequence in comparison to no insertion, and

ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and

iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector , and

iv) introducing said screening construct or screening vector with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and

v) identifying and/or selecting screening constructs or screening vectors with a changed readily detectable characteristic in comparison to no insertion, and

vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

One or more of the sequences to be assessed for their efficiency in transcription termination may be inserted into the screening vector or screening construct. In the case of insertion of two or more copies, in a preferred embodiment said DNA sequences to be assessed for their transcription termination efficiency are inserted into said insertion site in form of an inverted repeat. Thus, preferably the method for identification and isolation of transcription termination sequences comprises the steps of:

i) providing a screening construct or screening vector comprising in 5' to 3' direction

a) a promoter sequence, and

b) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic, and

c) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences,

ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and

iii) inserting at least two copies of a specific DNA sequence of said DNA sequences in form of an inverted repeat into said insertion site of said screening construct or screening vector, wherein insertion of an inverted repeat of an efficient transcription terminator into said insertion site allows expression of said additional sequences by said promoter sequence in comparison to no insertion, and

iv) introducing said screening constructs or screening vectors with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and

v) identifying and/or selecting screening constructs or screening vectors with said readily detectable characteristic in comparison to no insertion, and

vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

In another preferred embodiment of the invention, the method for identification and isolation of transcription termination sequences comprises the steps of:

i) providing a screening construct or screening vector comprising in 5' to 3' direction

a) a promoter sequence, and

b) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic, and embedded into said additional sequences one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences,

wherein insertion of an efficient transcription terminator into said insertion site suppresses full-length transcription of said additional sequences by said promoter sequence in comparison to no insertion, and

ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and

iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector, and

iv) introducing said screening constructs or screening vectors with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and

v) identifying and/or selecting screening constructs or screening vectors with a changed readily detectable characteristic in comparison to no insertion, and

vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

Preferably, the additional sequence is selected from the group consisting of positive selection marker, negative selection marker, counter selection marker, reporter genes, and toxic genes. In case of toxic genes, wherein said toxic gene may for example be a construct for gene silencing of an essential endogenous gene.

Preferably, the DNA sequence to be assessed for their transcription termination efficiency is provided by a method selected from the group consisting of:

i) provision of a selected sequence by amplification from a host genome, and

ii) provision of a library of sequences by fragmentation of a host genome.

More preferably, the DNA sequence to be assessed for their transcription termination efficiency is derived from a plant cell.

There various methods for insertion of said sequences into said insertion site. Preferably, the DNA sequences to be assessed for their transcription termination efficiency are inserted into said insertion site by a method selected from the group consisting of:

i) recombinational cloning, and

ii) insertion by sequence specific restriction and ligation.

## BRIEF DESCRIPTION OF THE DRAWINGS

**Fig. 1-A** **A1:** Graphic of preferred Method A: Insertion site (IS) for transcription terminator (TT) to be assessed is localized between promoter (P1) and additional sequences (AS), which are able to cause a readily detectable characteristic. In case of an efficient transcription terminator (+), transcription from the promoter P1 is stopped at said transcription terminator (symbolized by arrow below construct). No expression of the additional sequences occurs and no change in characteristic is caused (symbolized by crossed circle). In case of no efficient transcription termination (-), transcription from the promoter P1 read-through said alleged transcription terminator leading to expression of the additional sequences (symbolized by arrow below construct), thereby causing the change in the characteristic (symbolized by lightening symbol).

**A2:** Graphic of preferred Method A based on terminator as additional sequences. Insertion site (IS) for transcription terminator (TT) to be assessed is localized between promoter (P1) and additional sequences which in this case are constituted by an inverted repeat of a known transcription terminator (T). The second copy of said terminator (symbolized by upside letter) downstream of the promoter is in its functional orientation. In case of an efficient transcription terminator (+), transcription from the promoter P1 is stopped at said transcription terminator (symbolized by arrow below construct) and normal expression of the marker protein (M1) occurs (symbolized by black lightening symbol). In case of no efficient transcription termination (-), transcription from the promoter P1 read-through said alleged transcription terminator leading to expression of the inverted repeat of the known transcription terminator, causing gene silencing (GS) of the Marker M1 expression (symbolized by crossed circle). Preferably, the construct comprises a further expression cassette leading to expression of Marker M2, which functions as a positive control for general presence of the screening construct, bringing about a preferably different second phenotype (symbolized by white lightening symbol).

**Fig. 1-B** **A3:** Graphic of preferred Method A based on terminator as additional sequences. Insertion site (IS) for transcription terminator (TT) to be assessed is localized between two expression cassette for different marker genes, which are oriented head to head to eachother. The cassette for the marker M2 is terminated by a known transcription terminator. In case of an efficient termination of transcription by the inserted test terminator (+), transcription from the promoter P1 is stopped at said transcription terminator (symbolized by arrow below construct) and normal expression of the marker proteins (M1) and (M2) occurs (symbolized by black and white lightening symbols). In case of no efficient transcription termination (-), transcription from the promoter P1 read-through said alleged transcription terminator leading to expression of an RNA strand compemenatry to the one expressed from promoter P2, thereby causing hybridization of the transcript of the first marker gene (M1) with the constitutively expressed transcript from the second marker gene (M2). This causes a gene silencing (GS) of bothe Marker genes M1 and M2 (symbolized by crossed circle).

**A4:** Graphic of preferred Method A based on terminator as additional sequences. Insertion site (IS) for transcription terminator (TT) to be assessed is localized between promoter (P1) and additional sequences which in this case are constituted by non-transcribed DNA sequence preferably a fragment or a full length sequence of a known transcription terminator. The second expression cassette is oriented head to head to the first cassette and carries a second copy of said DNA sequence, terminator or terminator fragment. In case of an efficient termination of transcription by the inserted test terminator (+), transcription from the promoter P1 is stopped at said transcription terminator (symbolized by arrow below construct) and normal expression of the marker protein (M1) occurs (symbolized by black lightening symbol). In case of no efficient transcription termination (-), transcription from the promoter P1 readthrough said alleged transcription terminator leading to expression of the known transcription terminator or transcription terminator fragment, causing hybridization of the transcript of the first marker gene (M1) with the constitutively expressed transcript from the second marker gene (M2) which carries the identical 3´ UTR sequence, thereby causing a dose dependent repression of expression of both Marker genes M1 and M2 (symbolized by crossed circle). Preferably, the construct comprises a second marker gene (M2). As the effect of expression repression of both marker genes is dependent on the degree of hybridization between the two classes of transcripts it is possible to screen for intermediary phenotypes, allowing the selection of "weak" candidate terminator sequences or "tight" candidate terminator sequences e.g. by using different screening conditions.

**Fig. 2:** **B1:** Graphic of preferred Method B: Insertion site (IS) for transcription terminator (TT) to be assessed is downstream of the additional sequences (AS), which are able to cause a readily detectable characteristic. The transcription terminator (TT) to be assessed is inserted in form of an inverted repeat, wherein the first copy (symbolized by upside letter) downstream of the promoter is in its functional orientation. In case of an efficient transcription terminator (+), transcription from the promoter P1 is stopped at the first copy of the transcription terminator (symbolized by arrow below construct) and normal expression of the additional sequences (AS) occurs (symbolized by black lightening symbol). In case of no efficient transcription termination (-), transcription from the promoter P1 read-through both copies of said alleged transcription terminator leading to expression of the inverted repeat of said transcription terminator, causing gene silencing (GS) of the additional sequences (AS) expression (symbolized by crossed circle).

**B2:** Preferably, the construct comprises a further expression cassette, wherein AS2 is encoding for a different characteristic than AS1. Both expressions are silenced in case of an inefficient transcription terminator. In case of an efficient terminator, expression of both characteristics (symbolized by black and white lightening symbol, respectively) occurs.

**Fig. 3** Graphic of preferred Method C: Insertion site (IS) for transcription terminator (TT) to be assessed is localized in an intron (IN) localized in the additional sequences (AS), which are able to cause a readily detectable characteristic. In

case of an efficient transcription terminator (+), transcription from the promoter P1 is stopped at said transcription terminator (symbolized by arrow below construct). No full-length expression of the additional sequences occurs and no change in characteristic is caused (symbolized by crossed circle). In case of no efficient transcription termination (-), transcription from the promoter P1 read-through said alleged transcription terminator leading to full-length expression of the additional sequences (symbolized by arrow below construct), thereby causing the change in the characteristic (symbolized by lightening symbol).

**Fig. 4** Schematic presentation of the screening constructs:

**A:** Lo523 negative control construct: Binary vector corresponding to the screening construct without insertion of an additional transcription terminator sequence. Upon use of this construct for transformation of plant cells the *nptII* gene will be transcribed. As there is no functional terminator present downstream of this gene transcription proceeds through the nos terminator IR leading to a transcript with hairpin structure, which causes silencing of *nptII* gene expression. These cells cannot grow on selective medium containing Kanamycin. By visualization of the constitutively expressed GFP marker gene these non-growing cells can be distinguished from non-transformed cells.

**B:** Screening construct B: Binary vector containing a first expression cassette with a constitutively expressed reporter gene for selection of transformed from untransformed cells/plants followed by a second expression cassette containing a *nptII* selection marker driven by a strong constitutive promoter. Downstream of the *nptII* gene an IR of the *Agrobacterium* nos terminator sequence is inserted, consisting of a first repeat in antisense direction followed by a short spacer sequence derived from the GUS reporter gene and the second repeat of the nos terminator which is inserted in its functional 5´to 3´direction. The fragments to be tested for transcription terminator activity are to be inserted between the *nptII* gene and said nos terminator IR.

**D:** Lo546 positive control construct C: Binary vector derived from construct C where a long fragment of the rbCs E9 terminator sequence is inserted between *nptII* gene and nos terminator IR. The E9 terminator is believed to act as a highly efficient terminator and will therefore terminate transcription of the *nptII* gene resulting in normal expression levels of the selection marker which enable the growth of the transformed cells in presence of Kanamycin.

**E:** Lo239 negative control construct and base construct for insertion of candidate terminator sequences. Use of this binary construct for transformation leads to plants with constitutive expression of the *nptII* gene, producing transcripts carrying a defined 3´UTR. Upon seed development the expression of the ß-Glucuronidase marker is initiated by the seed-specific promotor. The resulting GUS transcripts carry the same defined 3´sequence as the *nptII* transcripts leading to hybridization between the two transcript classes and thereby causing loss of the phenotype associated with the two marker genes. The seeds cannot grow on selective medium containing Kanamycin and the GUS marker cannot be detected in the seeds.

**F:** Lo657 positive control construct. Binary vector derived from Lo239 by insertion of a long fragment of the rbCs E9 terminator between the GUS marker gene and the downstream sequences. By efficiently terminating the GUS transcripts which are produced within the cells of the developing seed the resulting transcript carry a different 3´sequence compared to the constitutively expressed *nptII* transcripts. Both marker cassettes are transcribed and translated without interferences leading to seedlings which are viable on kanamycin selection and allow detection of accumulated GUS protein.

**Fig. 5** pENTR construct – position 1 visual marker (reporter gene): Gateway Entry vector A containing the GFP reporter gene under control of a constitutive promoter. The vector is used in combination with the Gateway Entry vector B Lo376 (SEQ ID NO: 76) and the Gateway Entry vector C, Lo522a (Lo522b respectively), for recombination based construction of positive control constructs and in combination with Lo503a (Lo503b respectively) for recombination based construction of binary negative control constructs and the screening construct.

**Fig. 6** pENTR constructs for positive controls: The Gateway Entry vectors C contain the *nptII* selection marker gene under control of the constitutive STPT promoter. 3´to the selection marker a long fragment of the rbCs E9 terminator sequence is inserted in front of a nos terminator IR. The depicted constructs are used in combination with the Gateway Entry vectors A, Lo484, B Lo376 and C, Lo522 (Lo522b respectively), for recombination based construction of binary positive control constructs.

**Fig. 7** pENTR constructs for negative controls: The Gateway Entry vectors C contain the *nptII* selection marker gene under control of the constitutive STPT promoter. 3´to the selection marker a nos terminator IR is inserted. This results in transcription of the *nptII* Gene and the inverted sequence fragment of the nos terminator followed by the nos terminator sequence in functional 5´-3´orientation, which will by default cause a hairpin structure at the 3´end of the transcript. The depicted constructs are used in combination with the Gateway Entry vectors A, Lo484, B Lo376 and C, Lo522 (Lo522b respectively), for recombination based construction of binary positive control constructs.

**Fig. 8** pSUN1 constructs – negative controls: Binary construct derived from the recombination based insertion of the expression cassettes from the respective pENTR vectors into the Gateway Destination vector Lo442 pSUN1-R4R3 (SEQ ID NO: 77).

**Fig. 9** pSUN1 constructs – positive controls: Binary construct derived from the recombination based insertion of the expression cassettes from the respective pENTR vectors into the Gateway Destination vector Lo442 pSUN1-R4R3 (SEQ ID NO: 77).

**Fig.10** Binary constructs Lo239-pSUN3-GWs-B1-BnAK700::GUS::nosT-B2 (negative control) and Lo657- pSUN3-GWs-B1-BnAK700::GUS::E9::nosT::B2 (positive control) representing Fig4-E and F. Lo239 is derived from the Gateway based recombination of the pENTR Lo235 carrying the GUS marker cassette with the pSUN destination vector Lo125 pSUN3-GWs-NPTII carrying the *nptII* cassette. After modification of Lo235 by insertion of a long fragment of the rbCs E9 terminator downstream of the GUS marker gene the resulting Lo654 pENTR-BnAK700::GUS::E9::nosT is used for the Gateway based recombination with Lo125 pSUN3-GWs-NPTII to create the positive control construct Lo657- pSUN3-GWs-B1-BnAK700::GUS::E9::nosT:: B2

**Fig.11** Screening results of negative and positive control constructs: The constructs described in Fig.10 were used for floral dip transformation of Arabidopsis thaliana plants. The harvested seeds were tested for expression of the marker gene *nptII* by selection on Kanamycin. Whereas the seeds from plants which have been used for transformation with the positive control construct are viable on the selective medium, showing expression of the *nptII* gene (GUS gene expression is not shown but has been detected by X-Gluc reaction) the negative control construct yields only seeds which are not viable on Kanamycin and show no expression of the Gus marker gene.

**Fig. 12** Diagram of the constructs for identification of transcription terminators of interest (TOI).

(a) The construct for *in vitro* transcription assays. Gene 1 and 2 prefer to be the sequences that are not homologous to the plant genome.

(b-d) The constructs for *in vivo* assays. The regions indicated as "dsRNA" are the sequences that generate double-stranded RNA (dsRNA). These dsRNAs down-regulate an essential gene for plant cells (b), negative selection marker gene (SMG) (c), or a reporter gene such as fluorescence protein (FP)

(d) In the construct (d), the DNA downstream of TOI can be another reporter gene or any sequences.

(e) Screening system in yeast: The system allows efficient screening of random sequences. Control vectors contain NOS terminator, truncated NOS, no terminator, or a DNA fragment with unidentified sequence.

**Fig. 13** **A:** Control constructs (A-D) and a construct containing potential terminator candidate (E). GUS can be replaced with any reporter gene or a non-plant homologous DNA fragment including ATG and stop codon. In the control vectors, four variations are made: (A) 260 bp nopaline synthase terminator (NOS) including polyadenylation site, a cleavage sequence, and approximately 80 bp of nucleotides downstream of the cleavage sequence, (B) reverse orientation of NOS, (C) NOS DNA fragment either including polyadenylation signal and cleavage site or lacking the polyadenylation signal and downstream sequence, and (D) no terminator. Nopaline synthase terminator is replaced with various genomic fragments, which can be selected as potential terminator candidates. $Lu^F$ represents a fragment of luciferase gene (ap-

proximately 200 to 300 bp) as a read through region. $Lu^F$ can be replaced with a non-plant homologous DNA fragment (*e.g.* yeast intergenic sequences). Octopine synthase terminator is located in the end of the cassette to stabilize the transcripts including read through products. These constructs can be built in pUC based vector or a binary vector. TOI stands for terminator of interest. Although the NOS terminator has proven in the screening systems to be of only moderate efficiency, it can be used as kind of control terminator for the evaluation systems described herein.

**B:** An *in vitro* screening system. Two single strand fluorescence probes such as beacon probes that hybridize the read through region and polyadenylated RNA. The black bar represents polyadenylated RNA. The gray bar represents read through product. Probe 1 (black star) hybridizes the complementary sequence of the polyadenylated RNA. Probe 2 (gray star) hybridizes the complementary sequence of the read through region.

**C:** Control constructs (A, C) and constructs containing potential terminator candidates (B, D). RG represents any reporter gene. SMG represents any selectable marker gene. Approximately 260 bp Nopaline Synthase terminator (NOS) is used as a control terminator (A, C). Plant genomic fragment (<1 kb) is cloned between GUS or bar and dsRNA fragment to identify terminator of interest (TOI). Expression of dsRNA for essential gene (EG) causes lethal in plants due to down-regulation of essential gene in the transgenic plants (A, B). Expression of dsRNA for SMG causes lethal under specific selection pressure due to down-regulation of selectable marker gene in the transgenic plants (C, D). However, strong and tight terminator can limit the expression of dsRNA due to low levels or no read through resulting in producing transgenic plants. These constructs can be built in pUC based vector or a binary vector.

**D:** Control construct (A) is the vector into which control and test sequences were cloned. Positive control vector (B) comprises the NOS terminator sequence downstream of the GUS reporter gene. Negative control vector (C) comprises sequence obtained from an internal portion of a plant-expressible open-reading frame, and therefore should not possess transcriptional termination activity. Vector (D) represents vectors that comprise putative TOI candidates to be tested for terminator activity. These constructs were built into a pUC vector and used for transient analyses of TOI candidate sequences.

**E**: Control binary vectors (A, C) comprise no insertion of putative terminator sequences downstream of the primary reporter gene, GUS. Test vectors (B, D) comprise putative TOI candidates to be tested for transcriptional termination activity in stably transformed plants. Vectors (A, B) comprise a NOS terminator downstream of the secondary reporter gene, DsRed2; these constructs will be used to determine if efficiency of GUS termination by putative TOIs impacts expression of DsRed2. Vectors (C, D) contain no transcrip-

tional termination sequence for the secondary reporter gene DsRed2, and will be used to test for bidirectional transcriptional termination activity by TOI sequences that are juxtaposed between the 3' ends of the DsRed2 and GUS genes.

**Fig. 14** Maize leaf tissue following transient TOI assays. No GUS staining was observed in vectors that do not comprise a functional transcriptional terminator downstream of the GUS coding sequence (A & C). The presence of a functional terminator rescued GUS expression in the (+) control (B) vector as well as all four TOI candidate sequences (D-G).

**Fig. 15** Control construct (A) and a construct containing potential terminator candidates (B). The constructs are composed of strong constitutive promoter (*e.g.* maize ubiquitin promoter), FP1 (gene encoding fluorescent protein1), known (A: *e.g.* NOS) or novel (B) terminator, IRES (*e.g.* EMCV), FP2 (gene encoding fluorescent protein2), and octopine synthase terminator.

**Fig. 16** Terminator of interest (TOI) construct (A) and control constructs (B and C). A TOI is embedded within an intron of a lethal gene or a reporter gene (A). Control constructs will also be built without a TOI embedded in the intron (B) and with a known terminator, NOS, embedded in the intron (C). The lethal gene can be diphtheria toxin fragment A (DT-A) or any known lethal gene for plants in the art. The reporter gene can be green fluorescent protein or any known reporter gene functioning in plants in the art.

**Fig. 17** The modified PIV2 intron. PIV intron contains (1) a consensus 5' recognition sequence (2) high AU content after the 5' splice site, (3) high AU content before the 3' splice site, (4) a consensus 3' recognition sequence, (5) a consensus branchpoint sequence CURAY, (6) a polyU tract between the branchpoint sequence and the 3' splice site.

**Fig. 18** Terminator of interest (TOI) construct (1), and control constructs (2 and 3). Construct 4 expresses a dsRNA molecule which will target mRNA containing the RNAi target region of constructs 1 (leaky TOI) and construct 3 for degradation. ZmUbi (maize ubiquitin promoter), GFP (green fluorescence protein), OCS (Octopine synthase terminator), NOS (Nopaline synthase terminator), dsRNA (double-stranded RNA), siRNA (small interfering RNA), RNAi (RNA interference;silencing). For constructs 1, 2, and 3, the destination of the RNA produced is shown (translated or degraded) if construct 4 is also expressed in the same plant.

**Fig. 19** Schematic drawing of the inserts in vector pTOI3 (SEQ ID NO: 73) and pTOI4 (SEQ ID NO: 73)

**Fig. 20** A map of the pUC based expression vector that was used in transient analyses. Control and putative TOI sequences were cloned into the RsrII-SacI sites of this vector.

**Fig. 21** Maps of binary vectors to be used for analysis of TOI activity in stably-transformed plants. A – vectors comprise Nos terminator downstream of DsRed2 reporter gene, and will be used to determine if efficiency of GUS termination by putative TOIs affects expression of DsRed2. TOI sequences were inserted at the *Avr*II site. B – vectors comprise no terminator for DsRed2 and will be used to assess bidirectional transcriptional termination activity by TOI candidates. TOI sequences will be inserted at *Sac*I site.

## GENERAL DEFINITIONS

It is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, plant species or genera, constructs, and reagents described as such. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. In the description that follows, a number of terms used in recombinant DNA technology are utilized extensively. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

It must be noted that as used herein and in the appended claims, the singular forms "a," "and," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a vector" is a reference to one or more vectors and includes equivalents thereof known to those skilled in the art, and so forth.

The term "about" is used herein to mean approximately, roughly, around, or in the region of. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 20 percent up or down (higher or lower).

As used herein, the word "or" means any one member of a particular list and also includes any combination of members of that list.

The term “nucleotide” refers to a base-sugar-phosphate combination. Nucleotides are monomeric units of a nucleic acid sequence (DNA and RNA). The term nucleotide includes ribonucleoside triphosphatase ATP, UTP, CTG, GTP and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof Such derivatives include, for example, [αS]dATP, 7-deaza-dGTP and 7-deaza-dATP. The term nucleotide as used herein also refers to dideoxyribonucleoside triphosphates (ddNTPs) and their derivatives. Illustrated examples of dideoxyribonucleoside triphosphates include, but are not limited to, ddATP, ddCTP, ddGTP, ddITP, and ddTTP. According to the present invention, a "nucleotide" may be unlabeled or detectably labeled by well-known techniques. Detectable labels include, for example, radioactive isotopes, fluorescent labels, chemiluminescent labels, bioluminescent labels and enzyme labels.

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers or hybrids thereof in either single- or double-stranded, sense or antisense form.

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The term "nucleic acid" is used interchangeably herein with "gene", "cDNA, "mRNA", "oligonucleotide," and "polynucleotide".

The phrase "nucleic acid sequence" as used herein refers to a consecutive list of abbreviations, letters, characters or words, which represent nucleotides. In one embodiment, a nucleic acid can be a "probe" which is a relatively short nucleic acid, usually less than 100 nucleotides in length. Often a nucleic acid probe is from about 50 nucleotides in length to about 10 nucleotides in length. A "target region" of a nucleic acid is a portion of a nucleic acid that is identified to be of interest. A "coding region" of a nucleic acid is the portion of the nucleic acid which is transcribed and translated in a sequence-specific manner to produce into a particular polypeptide or protein when placed under the control of appropriate regulatory sequences. The coding region is said to encode such a polypeptide or protein.

The term "probe", as used herein, refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to a nucleotide sequence of interest. A probe may be single-stranded or double-stranded. It is contemplated that any probe used in the present invention will be labeled with any "reporter molecule," so that it is detectable in any detection system including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, calorimetric, gravimetric, magnetic, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The term “oligonucleotide” refers to a synthetic or natural molecule comprising a covalently linked sequence of nucleotides which are joined by a phosphodiester bond between the 3' position of the deoxyribose or ribose of one nucleotide and the 5' position of the deoxyribose or ribose of the adjacent nucleotide.

The term "sense" is understood to mean a nucleic acid having a sequence which is homologous or identical to a target sequence, for example a sequence which binds to a protein transcription factor and which is involved in the expression of a given gene. According to a preferred embodiment, the nucleic acid comprises a gene of interest and elements allowing the expression of the said gene of interest. The sense RNA can be employed for gene silencing in a co-suppression or sense-suppression gene silencing approach. Expression of genes that when transcribed produce RNA transcripts that are identical or at least very similar to transcripts of endogenous genes can mediate gene silencing in an as yet not fully understood way of inhibition of gene expression referred to as co-suppression (disclosed by Napoli 1990; Jorgensen 1996; Goring 1991; Smith 1990; Van der Krol 1990). The expressed RNA can represent the endogenous target entirely or in part. Translation is nor required, transcription is sufficient. Application in plants is described (Napoli 1990; US 5,034,323).

The term "antisense" is understood to mean a nucleic acid having a sequence complementary to a target sequence, for example a messenger RNA (mRNA) sequence the blocking of whose expression is sought to be initiated by hybridization with the tar-

get sequence. To maximize the antisense effects in a plant host, the use of homologous genes is preferred. With homologous is meant obtainable from the same plant species as the plant host. Heterologous, for the purpose of this specification shall mean obtainable from a different plant or non-plant species. Heterologous shall also comprise synthetic analogs of genes, modified in their mRNA encoding nucleic acid sequence to diverge at least 5% of the host gene. Gene silencing by antisense RNA is numerously described in the art (including various applications in plants; *e.g.* Sheehy 1988; US 4,801,340; Mol 1990). A variation of the antisense approach is the use of α-anomeric nucleic acid sequences. Such α-anomeric sequences are forming specific double-stranded hybrids with complementary RNA, wherein in contrast to "normal" antisense RNA (or β-nucleic acids) both strands are in parallel to each other (Gautier 1987).

The term "dsRNAi" or "double-stranded RNA interference" is intended to mean the method of gene silencing by expression of a RNA molecule corresponding to an endogenous gene together with its complementary RNA strand, thus providing two RNA sequences which may form by hybridization a double-stranded RNA structure. The two RNA strands may be on separate molecules or may be part of one molecule, thus forming a so-called self-complementary hairpin structure. Self-complementary hairpin forming RNA structure may be expressed for example from a DNA comprising an "inverted repeat" of a double-stranded DNA fragment. In this context the term "inverted repeat" is intended to mean the orientation of two fragments of double-stranded DNA (which are substantially identical or – preferably identical in sequence) in one double stranded DNA molecule in an inverted orientation (*i.e.* in a "head" to "head" or "tail" to tail" orientation so that the sense-strand of the first fragment is fused to the antisense strand of the second and *vice versa*). Preferably, the hairpin forming dsRNA may include a linker (*e.g.*, an intron sequence for example the intron of the ST-LS1 gene from potato; Vancanneyt 1990) connecting the two complementary strands (*e.g.*, as described in WO 99/53050). The method of dsRNAi is well described in the art for various organisms including animal and plant organism (*e.g.,* Matzke 2000; Fire 1998; WO 99/32619; WO 99/53050; WO 00/68374; WO 00/44914; WO 00/44895; WO 00/49035; WO 00/63364). These references and the methods disclosed therein are explicitly incorporated by reference. The phenotype of a dsRNAi expressing cell or organism is similar to that of a knock-out mutant, often resulting in complete gene silencing (Waterhouse 1998). The term "double-stranded RNA" or "dsRNA" as used herein is intended to mean one or more ribonucleic acid sequences, which because of complementary sequences are theoretically (*i.e.* according to the base-pairing rules of Watson and Crick) and/or practically (*e.g.*, because of hybridization experiments *in vitro* and/or *in vivo*) capable to form double-stranded RNA structures. The person skilled in the art is aware of the fact that formation of a double-stranded RNA structure is an equilibrium between single-stranded and double-stranded forms. Preferably, the relation between double-stranded (*i.e.* hybridized) and single-stranded (*i.e.* non-hybridized or dissociated) forms is at least 1:10, preferably at least 1:1, more preferably at least 10:1. One strand of the dsRNA is essentially identical to the sequence of the endogenous gene. Essentially identical in this context means that a 100% identity is not required for efficient gene silencing, but that the dsRNA sequence may comprise insertions, deletions and point mutations in comparison to the target sequence. Preferably the homology between the dsRNA sequence and at least part of the target sequence is at least 60%, preferably at least 80%, more preferably at least 90%, most preferably 100%. Alternatively, an essential identity is one which allows for hybridization of the two sequences

under high stringency conditions. The part of the target sequence which is having the homology with the dsRNA has a length of at least 23 bases, preferably at least 50 bases, more preferably at least 100 bases. Said part of the target gene may resemble various part of the gene, but is preferably part which encodes for the mRNA sequence transcribed from said gene.

The term "ribozyme" is intended to mean catalytic RNA-molecules, which are capable to induce sequence-specific cleavage of a target RNA (Tanner 1999). Preparation and use of ribozymes is disclosed in the art (Haseloff 1988; Haselhoff & Gerlach 1988; Steinecke 1992; de Feyter R 1996). Preferred are "hammerhead"-ribozymes (Haselhoff & Gerlach 1988). Disclosed are methods for gene silencing based on customized ribozymes (EP 0 291 533, EP 0 321 201, EP 0 360 257). Use in plants and plant cells is also disclosed (Steinecke 1992; de Feyter 1996). Suitable target sequences are ribozymes can be derived as described (Steinecke 1995) by secondary structure calculation of ribozyme and target sequences and the interaction thereof (Bayley 1992; Lloyd 1994). For example derivatives of the Tetrahymena L-19 IVS RNA can be employed and adapted to virtually any target sequence (US 4,987,071; US 5,116,742). Alternatively, ribozymes can be selected by screening of diversified ribozyme libraries (Bartel 1993).

The term "gene" refers to a coding region operably joined to appropriate regulatory sequences capable of regulating the expression of the polypeptide in some manner. A gene includes untranslated regulatory regions of DNA (*e.g.*, promoters, enhancers, repressors, *etc.*) preceding (upstream) and following (downstream) the coding region (open reading frame, ORF) as well as, where applicable, intervening sequences (*i.e.*, introns) between individual coding regions (*i.e.*, exons).

As used herein the term "coding region" when used in reference to a structural gene refers to the nucleotide sequences which encode the amino acids found in the nascent polypeptide as a result of translation of a mRNA molecule. The coding region is bounded, in eukaryotes, on the 5'-side by the nucleotide triplet "ATG" which encodes the initiator methionine and on the 3'-side by one of the three triplets which specify stop codons (*i.e.*, TAA, TAG, TGA). In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5'- and 3'-end of the sequences, which are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5'-flanking region may contain regulatory sequences such as promoters and enhancers, which control or influence the transcription of the gene. The 3'-flanking region may contain sequences, which direct the termination of transcription, posttranscriptional cleavage and polyadenylation.

The term "amplification" refers to any *in vitro* method for increasing a number of copies of a nucleotide sequence with the use of a polymerase. Nucleic acid amplification results in the incorporation of nucleotides into a DNA and/or RNA molecule or primer thereby forming a new molecule complementary to a template. The formed nucleic acid molecule and its template can be used as templates to synthesize additional nucleic acid molecules. As used herein, one amplification reaction may consist of many rounds of replication. DNA amplification reactions include, for example, polymerase chain re-

action (PCR). One PCR reaction may consist of 5 to 100 "cycles" of denaturation and synthesis of a DNA molecule.

The terms "polypeptide", "peptide", "oligopeptide", "polypeptide", "gene product", "expression product" and "protein" are used interchangeably herein to refer to a polymer or oligomer of consecutive amino acid residues.

The term "isolated" as used herein means that a material has been removed from its original environment. For example, a naturally occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides can be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and would be isolated in that such a vector or composition is not part of its original environment.

Preferably, the term "isolated" when used in relation to a nucleic acid, as in "an isolated nucleic acid sequence" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature. For example, a given DNA sequence (*e.g.,* a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. For example, an isolated nucleic acid sequence encoding for a specific trait includes, by way of example, such nucleic acid sequences in cells which ordinarily contain said nucleic acid sequence, wherein said nucleic acid sequence is in a chromosomal or extrachromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid sequence may be present in single-stranded or double-stranded form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain at a minimum at least a portion of the sense or coding strand (*i.e.,* the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and antisense strands (*i.e.,* the nucleic acid sequence may be double-stranded).

As used herein, the term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An "isolated nucleic acid sequence" is therefore a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated.

As used herein, the terms "complementary" or "complementarity" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with an-

other base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

A "complement" of a nucleic acid sequence as used herein refers to a nucleotide sequence whose nucleic acids show total complementarity to the nucleic acids of the nucleic acid sequence.

The term "wild-type", "natural" or of "natural origin" means with respect to an organism, polypeptide, or nucleic acid sequence, that said organism is naturally occurring or available in at least one naturally occurring organism which is not changed, mutated, or otherwise manipulated by man.

The term "transgenic" or "recombinant" when used in reference to a cell refers to a cell which contains a transgene, or whose genome has been altered by the introduction of a transgene. The term "transgenic" when used in reference to a tissue or to a plant refers to a tissue or plant, respectively, which comprises one or more cells that contain a transgene, or whose genome has been altered by the introduction of a transgene. Transgenic cells, tissues and plants may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into a target cell or integration of the transgene into a chromosome of a target cell by way of human intervention, such as by the methods described herein.

The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations. A transgene may be an "endogenous DNA sequence," or a "heterologous DNA sequence" (*i.e.,* "foreign DNA"). The term "endogenous DNA sequence" refers to a nucleotide sequence which is naturally found in the cell into which it is introduced so long as it does not contain some modification (*e.g.*, a point mutation, the presence of a selectable marker gene, *etc.*) relative to the naturally-occurring sequence. The term "heterologous DNA sequence" refers to a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Heterologous DNA also includes an endogenous DNA sequence which contains some modification. Generally, although not necessarily, heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, selectable marker proteins (*e.g.*, proteins which confer drug resistance), *etc.* Preferably, the term “transgenic” or "recombinant" with respect to a regulatory sequence (*e.g.*, a promoter of the invention) means that said regulatory sequence is covalently joined and adjacent to a nucleic acid to which it is not adjacent in its natural environment.

The term "foreign gene" refers to any nucleic acid (*e.g.*, gene sequence) which is introduced into the genome of a cell by experimental manipulations and may include gene sequences found in that cell so long as the introduced gene contains some modification (*e.g.*, a point mutation, the presence of a selectable marker gene, *etc.*) relative to the naturally-occurring gene.

"Recombinant polypeptides" or "recombinant proteins" refer to polypeptides or proteins produced by recombinant DNA techniques, *i.e.*, produced from cells transformed by an exogenous recombinant DNA construct encoding the desired polypeptide or protein. Recombinant nucleic acids and polypeptide may also comprise molecules which as such does not exist in nature but are modified, changed, mutated or otherwise manipulated by man.

The terms "heterologous nucleic acid sequence" or "heterologous DNA" are used interchangeably to refer to a nucleotide sequence which is ligated to a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Generally, although not necessarily, such heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed.

The terms "organism", "host", "target organism" or "host organism" are referring to any prokaryotic or eukaryotic organism that can be a recipient of the screening construct or screening vector. A "host," as the term is used herein, includes prokaryotic or eukaryotic organisms that can be genetically engineered. For examples of such hosts, see Maniatis 1989. Included are entire organisms but also organs, parts, cells, cultures, and propagatable material derived therefrom. Preferred are microorganisms, non-human animal and plant organisms. Preferred microorganisms are bacteria, yeasts, algae or fungi.

Preferred bacteria are bacteria of the genus *Escherichia, Corynebacterium, Bacillus, Erwinia, Agrobacterium, Flavobacterium, Alcaligenes, Clostridium, Proionibacterium, Butyrivibrio, Eubacterium, Lactobacillus, Phaeodactylum, Colpidium, Mortierella, Entomophthora, Mucor, Crypthecodinium or cyanobacteria*, for example of the genus *Synechocystis*. Especially preferred are microorganisms which are capable of infecting plants and thus of transferring the constructs according to the invention. Preferred microorganisms are those from the genus *Agrobacterium* and, in particular, the species *Agrobacterium tumefaciens*.

Preferred yeasts are *Candida, Saccharomyces, Hansenula or Pichia*. Preferred fungi are *Aspergillus, Trichoderma, Ashbya, Neurospora, Fusarium, Beauveria* or other fungi. Plant organisms are furthermore, for the purposes of the invention, other organisms which are capable of photosynthetic activity such as, for example, algae or cyanobacteria, and also mosses. Preferred algae are green algae such as, for example, algae of the genus *Haematococcus, Phaedactylum tricornatum, Volvox* or *Dunaliella*.

Preferred eukaryotic cells and organism comprise plant cells and organisms, animal cells, and non-human animal organism, including eukaryotic microorganism such as yeast, algae, or fungi.

"Non-human animal organisms" includes but is not limited to non-human vertebrates and invertebrates. Preferred are fish species, non-human mammals such as cow, horse, sheep, goat, mouse, rat or pig, birds such as chicken or goose. Preferred animal cells comprise for example CHO, COS, HEK293 cells. Invertebrate organisms include

for example nematodes and insects. Insect cells include for example *Drosophila* S2 and *Spodoptera* Sf9 or Sf21 cells.

Preferred nematodes are those which are capable to invade plant, animal or human organism. Preferred namtodes include for example nematodes of the genus *Ancylostoma, Ascaridia, Ascaris, Bunostomum, Caenorhabditis, Capillaria, Chabertia, Cooperia, Dictyocaulus, Haemonchus, Heterakis, Nematodirus, Oesophagostomum, Ostertagia, Oxyuris, Parascaris, Strongylus, Toxascaris, Trichuris, Trichostrongylus, Tfhchonema, Toxocara* or *Uncinaria*. Especially preferred are plant parasitic nematodes such as *Bursaphalenchus, Criconemella, Diiylenchus, Ditylenchus, Globodera, Helicotylenchus, Heterodera, Longidorus, Melodoigyne, Nacobbus, Paratylenchus, Pratylenchus, Radopholus, Rotelynchus, Tylenchus* or *Xiphinema*. Preferred insects comprise those of the genus *Coleoptera, Diptera, Lepidoptera*, and *Homoptera*.

Preferred fungi are *Aspergillus, Trichoderma, Ashbya, Neurospora, Fusarium, Beauveria* or other fungi described in Indian Chem Engr. Section B. Vol 37, No 1,2 (1995) on page 15, table 6. Especially preferred is the filamentic Hemiascomycete *Ashbya gossypii.*

Preferred yeasts are *Candida, Saccharomyces, Hansenula* or *Pichia*, especially preferred are *Saccharomyces cerevisiae* and *Pichia pastoris* (ATCC Accession No. 201178).

The term "plant" or "plant organism" as used herein refers to a plurality of plant cells which are largely differentiated into a structure that is present at any stage of a plant's development. Such structures include one or more plant organs including, but are not limited to, fruit, shoot, stem, leaf, flower petal, *etc.* Host or target organisms which are preferred as transgenic organisms are especially plants. Included within the scope of the invention are all genera and species of higher and lower plants of the plant kingdom. Included are furthermore the mature plants, seeds, shoots and seedlings and parts, propagation material and cultures derived therefrom, for example cell cultures. The term "mature plants" is understood as meaning plants at any developmental stage beyond the seedling. The term "seedling" is understood as meaning a young, immature plant in an early developmental stage.

Annual, biennial, monocotyledonous and dicotyledonous plants are preferred host organisms for the generation of transgenic plants. The expression of genes is furthermore advantageous in all ornamental plants, useful or ornamental trees, flowers, cut flowers, shrubs or lawns. Plants which may be mentioned by way of example but not by limitation are angiosperms, bryophytes such as, for example, *Hepaticae* (liverworts) and *Musci* (mosses); *Pteridophytes* such as ferns, horsetail and club mosses; gymnosperms such as conifers, cycads, ginkgo and *Gnetatae*; algae such as *Chlorophyceae, Phaeophpyceae, Rhodophyceae, Myxophyceae, Xanthophyceae, Bacillariophyceae* (diatoms) and *Euglenophyceae*.

Preferred are plants which are used for food or feed purpose such as the families of the *Leguminosae* such as pea, alfalfa and soya; *Gramineae* such as rice, maize, wheat, barley, sorghum, millet, rye, triticale, or oats; the family of the *Umbelliferae*, especially the genus *Daucus*, very especially the species *carota* (carrot) and *Apium*, very espe-

cially the species *Graveolens dulce* (celery) and many others; the family of the *Solanaceae*, especially the genus *Lycopersicon*, very especially the species *esculentum* (tomato) and the genus *Solanum*, very especially the species *tuberosum* (potato) and *melongena* (egg plant), and many others (such as tobacco); and the genus *Capsicum*, very especially the species *annuum* (peppers) and many others; the family of the *Leguminosae*, especially the genus *Glycine*, very especially the species *max* (soybean), alfalfa, pea, lucerne, beans or peanut and many others; and the family of the *Cruciferae* (*Brassicacae*), especially the genus *Brassica*, very especially the species *napus* (oil seed rape), *campestris* (beet), *oleracea* cv Tastie (cabbage), *oleracea* cv Snowball Y (cauliflower) and *oleracea* cv Emperor (broccoli); and of the genus *Arabidopsis*, very especially the species *thaliana* and many others; the family of the *Compositae*, especially the genus *Lactuca*, very especially the species *sativa* (lettuce) and many others; the family of the *Asteraceae* such as sunflower, Tagetes, lettuce or Calendula and many other; the family of the *Cucurbitaceae* such as melon, pumpkin/squash or zucchini, and linseed. Further preferred are cotton, sugar cane, hemp, flax, chillies, and the various tree, nut and wine species. Very especially preferred are *Arabidopsis thaliana*, *Nicotiana tabacum*, *Tagetes erecta*, *Calendula officinalis*, *Gycine max*, *Zea mays*, *Oryza sativa, Triticum aestivum, Pisum sativum, Phaseolus vulgaris, Hordium vulgare, Brassica napus*.

The term "cell" refers to a single cell. The term "cells" refers to a population of cells. The population may be a pure population comprising one cell type. Likewise, the population may comprise more than one cell type. In the present invention, there is no limit on the number of cell types that a cell population may comprise. The cells may be synchronized or not synchronized, preferably the cells are synchronized.

The term "organ" with respect to a plant (or "plant organ") means parts of a plant and may include (but shall not limited to) for example roots, fruits, shoots, stem, leaves, anthers, sepals, petals, pollen, seeds, *etc.*

The term "tissue" with respect to a plant (or "plant tissue") means arrangement of multiple plant cells including differentiated and undifferentiated tissues of plants. Plant tissues may constitute part of a plant organ (*e.g.,* the epidermis of a plant leaf) but may also constitute tumor tissues and various types of cells in culture (*e.g.*, single cells, protoplasts, embryos, calli, protocorm-like bodies, *etc.*). Plant tissue may be *in planta*, in organ culture, tissue culture, or cell culture.

The term "chromosomal DNA" or "chromosomal DNA-sequence" is to be understood as the genomic DNA of the cellular nucleus independent from the cell cycle status. Chromosomal DNA might therefore be organized in chromosomes or chromatids, they might be condensed or uncoiled. An insertion into the chromosomal DNA can be demonstrated and analyzed by various methods known in the art like *e.g.*, polymerase chain reaction (PCR) analysis, Southern blot analysis, fluorescence *in situ* hybridization (FISH), and *in situ* PCR.

The term "structural gene" as used herein is intended to mean a DNA sequence that is transcribed into mRNA which is then translated into a sequence of amino acids characteristic of a specific polypeptide.

The term "nucleotide sequence of interest" refers to any nucleotide sequence, the manipulation of which may be deemed desirable for any reason (*e.g.*, confer improved qualities), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (*e.g.*, reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factors, *etc.*), and non-coding regulatory sequences which do not encode an mRNA or protein product, (*e.g.*, promoter sequence, polyadenylation sequence, termination sequence, enhancer sequence, *etc.*).

The term "expression" refers to the biosynthesis of a gene product. For example, in the case of a structural gene, expression involves transcription of the structural gene into mRNA and - optionally - the subsequent translation of mRNA into one or more polypeptides.

The term "expression cassette" or "expression construct" as used herein is intended to mean the combination of any nucleic acid sequence to be expressed in operable linkage with a promoter sequence and - optionally – additional elements (like *e.g.,* terminator and/or polyadenylation sequences) which facilitate expression of said nucleic acid sequence.

The term "promoter," "promoter element," or "promoter sequence" as used herein, refers to a DNA sequence which when ligated to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. A promoter is typically, though not necessarily, located 5' (*i.e.*, upstream) of a nucleotide sequence of interest (*e.g.*, proximal to the transcriptional start site of a structural gene) whose transcription into mRNA it controls, and provides a site for specific binding by RNA polymerase and other transcription factors for initiation of transcription. A repressible promoter's rate of transcription decreases in response to a repressing agent. An inducible promoter's rate of transcription increases in response to an inducing agent. A constitutive promoter's rate of transcription is not specifically regulated, though it can vary under the influence of general metabolic conditions.

Promoters may be tissue specific or cell specific. The term "tissue specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (*e.g.*, petals) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (*e.g.*, roots). Tissue specificity of a promoter may be evaluated by, for example, operably linking a reporter gene to the promoter sequence to generate a reporter construct, introducing the reporter construct into the genome of a plant such that the reporter construct is integrated into every tissue of the resulting transgenic plant, and detecting the expression of the reporter gene (*e.g.*, detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic plant. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the promoter is specific for the tissues in which greater levels of expression are detected. The term "cell type specific" as applied to a promoter refers to a promoter which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell

type specific" when applied to a promoter also means a promoter capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue. Cell type specificity of a promoter may be assessed using methods well known in the art, *e.g.*, GUS activity staining or immunohistochemical staining. Briefly, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is controlled by the promoter. A labeled (*e.g.*, peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (*e.g.*, with avidin/biotin) by microscopy.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (*e.g.*, heat shock, chemicals, light, *etc.*). Typically, constitutive promoters are capable of directing expression of a transgene in substantially any cell and any tissue. In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nuclei acid sequence in the presence of a stimulus (*e.g.*, heat shock, chemicals, light, *etc.*) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The term “transcription terminator” or “transcription terminator sequence” as used herein is intended to mean a sequence which leads to or initiates a stop of transcription of a nucleic acid sequence initiated from a promoter. Preferably, a transcription terminator sequences is furthermore comprising sequences which cause polyadenylation of the transcript. A transcription terminator may, for example, comprise one or more polyadenylation signal sequences, one or more polyadenylation attachment sequences, and downstream sequence of various lengths which causes termination of transcription. It has to be understood that also sequences downstream of sequences coding for the 3’-untranslated region of an expressed RNA transcript may be part of a transcription terminator although the sequence itself is not expressed as part of the RNA transcript. Furthermore, a transcription terminator may comprise additional sequences, which may influence its functionality, such a 3’-untranslated sequences (*i.e.* sequences of a gene following the stop-codon of the coding sequence). Transcription termination may involve various mechanisms including but not limited to induced dissociation of RNA polymerase II from their DNA template. As virtually all biological reactions transcription termination is never of 100% efficiency. The term “transcription termination efficiency” or “efficiency or transcription termination” as used herein is indicating the ratio between the frequencies of stops (or termination) of transcription in the region of said transcription terminator to the frequency of read-through transcription beyond said transcription terminator. The term “tight” or “efficient” in relation to transcription termination sequence as used herein is understood as a transcription termination sequence for which the efficiency of transcription termination is at least 10 (*i.e.* stop/read-through ratio of 10:1), preferably at least 100 (*i.e.* stop/read-through ratio of 100:1), more preferably 1000 (*i.e.* stop/read-through ratio of 1000:1). Transcription may end at one or more specific base pairs within said transcription terminator sequence. In consequence, there might be variability in the length of the transcript. However, preferably transcription termination has a low variability and end to at least 50%, preferably at least 80%, more preferably at least 90% at one specific base pair as judged by the

resulting transcript length (excluding the poly-A tail).

The term "operable linkage" or "operably linked" is to be understood as meaning, for example, the sequential arrangement of a regulatory element (*e.g.* a promoter) with a nucleic acid sequence to be expressed and, if appropriate, further regulatory elements (such as *e.g.*, a terminator) in such a way that each of the regulatory elements can fulfill its intended function to allow, modify, facilitate or otherwise influence expression of said nucleic acid sequence. The expression may result depending on the arrangement of the nucleic acid sequences in relation to sense or antisense RNA. To this end, direct linkage in the chemical sense is not necessarily required. Genetic control sequences such as, for example, enhancer sequences, can also exert their function on the target sequence from positions which are further away, or indeed from other DNA molecules. Preferred arrangements are those in which the nucleic acid sequence to be expressed recombinantly is positioned behind the sequence acting as promoter, so that the two sequences are linked covalently to each other. The distance between the promoter sequence and the nucleic acid sequence to be expressed recombinantly is preferably less than 200 base pairs, especially preferably less than 100 base pairs, very especially preferably less than 50 base pairs. Operable linkage, and an expression construct, can be generated by means of customary recombination and cloning techniques as described (*e.g.*, in Maniatis 1989; Silhavy 1984; Ausubel 1987; Gelvin 1990). However, further sequences which, for example, act as a linker with specific cleavage sites for restriction enzymes, or as a signal peptide, may also be positioned between the two sequences. The insertion of sequences may also lead to the expression of fusion proteins. Preferably, the expression construct, consisting of a linkage of promoter and nucleic acid sequence to be expressed, can exist in a vector-integrated form and be inserted into a plant genome, for example by transformation.

The term "transformation" as used herein refers to the introduction of genetic material (*e.g.*, a transgene) into a cell. Transformation of a cell may be stable or transient. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) which detects the presence of a polypeptide encoded by one or more of the transgenes. Alternatively, transient transformation may be detected by detecting the activity of the protein (*e.g.*, β-glucuronidase) encoded by the transgene (*e.g.*, the *uid A* gene) as demonstrated herein [*e.g.*, histochemical assay of GUS enzyme activity by staining with X-Gluc which gives a blue precipitate in the presence of the GUS enzyme; and a chemiluminescent assay of GUS enzyme activity using the GUS-Light kit (Tropix)]. The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes.

In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell, preferably resulting in chromosomal integration and stable heritability through meiosis. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell which has stably inte-

grated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene. Transformation also includes introduction of genetic material into plant cells in the form of plant viral vectors involving epichromosomal replication and gene expression which may exhibit variable properties with respect to meiotic stability. Stable transformation also includes introduction of genetic material into cells in the form of viral vectors involving epichromosomal replication and gene expression which may exhibit variable properties with respect to meiotic stability.

Cloning and transformation techniques for manipulation of ciliates and algae are well known in the art (WO 98/01572; Falciatore 1999; Dunahay 1995).

Principally speaking transformation techniques suitable for plant cells or organisms (as described below) can also be employed for animal or yeast organism and cells. Preferred are direct transformation techniques such as calcium phosphate or liposome mediated transformation, or electroporation.

The terms "infecting" and "infection" with a bacterium refer to co-incubation of a target biological sample, (*e.g.*, cell, tissue, *etc.*) with the bacterium under conditions such that nucleic acid sequences contained within the bacterium are introduced into one or more cells of the target biological sample.

The term "*Agrobacterium*" refers to a soil-borne, Gram-negative, rod-shaped phytopathogenic bacterium which causes crown gall. The term "*Agrobacterium*" includes, but is not limited to, the strains *Agrobacterium tumefaciens*, (which typically causes crown gall in infected plants), and *Agrobacterium rhizogenes* (which causes hairy root disease in infected host plants). Infection of a plant cell with *Agrobacterium* generally results in the production of opines (*e.g.*, nopaline, agropine, octopine *etc.*) by the infected cell. Thus, *Agrobacterium* strains which cause production of nopaline (*e.g.*, strain LBA4301, C58, A208) are referred to as "nopaline-type" *Agrobacteria*; *Agrobacterium* strains which cause production of octopine (*e.g.*, strain LBA4404, Ach5, B6) are referred to as "octopine-type" *Agrobacteria*; and *Agrobacterium* strains which cause production of agropine (*e.g.*, strain EHA105, EHA101, A281) are referred to as "agropine-type" *Agrobacteria*.

The terms "bombarding, "bombardment," and "biolistic bombardment" refer to the process of accelerating particles towards a target biological sample (*e.g.*, cell, tissue, *etc.*) to effect wounding of the cell membrane of a cell in the target biological sample and/or entry of the particles into the target biological sample. Methods for biolistic bombardment are known in the art (*e.g.*, US 5,584,807, the contents of which are herein incorporated by reference), and are commercially available (*e.g.*, the helium gas-driven microprojectile accelerator (PDS-1000/He) (BioRad).

The terms "homology" or "identity" when used in relation to nucleic acids refers to a degree of complementarity. Homology or identity between two nucleic acids is understood as meaning the identity of the nucleic acid sequence over in each case the entire length of the sequence, which is calculated by comparison with the aid of the program algorithm GAP (Wisconsin Package Version 10.0, University of Wisconsin, Genetics

Computer Group (GCG), Madison, USA) with the parameters being set as follows:

Gap Weight: 12 Length Weight: 4

Average Match: 2,912 Average Mismatch:-2,003

Alternatively, a partially complementary sequence is understood to be one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (*i.e.*, an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous sequence to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described infra. When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize to the single-stranded nucleic acid sequence under conditions of low stringency as described infra.

The terms "hybridization" and "hybridizing" as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing." (Coombs 1994). Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the Tm of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein, the term "Tm" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the Tm of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the Tm value may be calculated by the equation: Tm=81.5+0.41(% G+C), when a nucleic acid is in aqueous solution at 1 M NaCl (see *e.g.*, Anderson 1985). Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of Tm.

Low stringency conditions when used in reference to nucleic acid hybridization com-

prise conditions equivalent to binding or hybridization at 68°C. in a solution consisting of 5x SSPE (43.8 g/L NaCl, 6.9 g/L $NaH_2PO_4.H_2O$ and 1.85 g/L EDTA, pH adjusted to 7.4 with NaOH), 1% SDS, 5x Denhardt's reagent [50x Denhardt's contains the following per 500 mL: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/mL denatured salmon sperm DNA followed by washing in a solution comprising 0.2x SSPE, and 0.1% SDS at room temperature when a DNA probe of about 100 to about 1000 nucleotides in length is employed.

High stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C. in a solution consisting of 5x SSPE, 1% SDS, 5x Denhardt's reagent and 100 µg/mL denatured salmon sperm DNA followed by washing in a solution comprising 0.1x SSPE, and 0.1% SDS at 68°C. when a probe of about 100 to about 1000 nucleotides in length is employed.

The term "equivalent" when made in reference to a hybridization condition as it relates to a hybridization condition of interest means that the hybridization condition and the hybridization condition of interest result in hybridization of nucleic acid sequences which have the same range of percent (%) homology. For example, if a hybridization condition of interest results in hybridization of a first nucleic acid sequence with other nucleic acid sequences that have from 80% to 90% homology to the first nucleic acid sequence, then another hybridization condition is said to be equivalent to the hybridization condition of interest if this other hybridization condition also results in hybridization of the first nucleic acid sequence with the other nucleic acid sequences that have from 80% to 90% homology to the first nucleic acid sequence.

When used in reference to nucleic acid hybridization the art knows well that numerous equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, *etc.*) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency hybridization different from, but equivalent to, the above-listed conditions. Those skilled in the art know that whereas higher stringencies may be preferred to reduce or eliminate non-specific binding between the nucleotide sequence of interest and other nucleic acid sequences, lower stringencies may be preferred to detect a larger number of nucleic acid sequences having different homologies to the nucleotide sequence of interest.

The term "recognition sequence" refers to a particular sequences which a protein, chemical compound, DNA, or RNA molecule (*e.g.*, restriction endonuclease, a modification methylase, or a recombinase) recognizes and binds. With respect to a recombinase a recognition sequence will usually refer to a recombination site. For example, the recognition sequence for Cre recombinase is loxP which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8 base pair core sequence (see Sauer 1994; figure 1). Other examples of recognition sequences are the attB, attP, attL, and attR sequences which are recognized by the recombinase enzyme λ Integrase. attB is an approximately 25 base pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair

overlap region. attP is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins integration host factor (IHF), FIS and excisionase (Xis) (see Landy 1993). Such sites may also be engineered according to the present invention to enhance production of products in the methods of the invention. When such engineered sites lack the P1 or H1 domains to make the recombination reactions irreversible (*e.g.*, attR or attP), such sites may be designated attR' or attP' to show that the domains of these sites have been modified in some way.

The term "recombinase" is referring to an enzyme which catalyzes the exchange of DNA segments at specific recombination sites.

The term "recombinational cloning" is referring to a method, whereby segments of nucleic acid molecules or populations of such molecules are exchanged, inserted, replaced, substituted or modified, *in vitro* or *in vivo*, by action of a site-specific recombinase.

The term "Recombination proteins" refers to polypeptide including excisive or integrative proteins, enzymes, co-factors or associated proteins that are involved in recombination reactions involving one or more recombination sites (Landy 1993).

Repression cassette: is a nucleic acid segment that contains a repressor of a Selectable marker present in the subcloning vector.

The term "site-specific recombinase" as used herein is referring to a type of recombinase which typically has at least the following four activities (or combinations thereof): (1) recognition of one or two specific nucleic acid sequences; (2) cleavage of said sequence or sequences; (3) topoisomerase activity involved in strand exchange; and (4) ligase activity to reseal the cleaved strands of nucleic acid (see Sauer 1994). Conservative site-specific recombination is distinguished from homologous recombination and transposition by a high degree of specificity for both partners. The strand exchange mechanism involves the cleavage and rejoining of specific DNA sequences in the absence of DNA synthesis (Landy 1989).

The term "vector" is referring to a nucleic acid molecule (preferably DNA) that provides a useful biological or biochemical property to an inserted nucleic acid sequence, preferably allows replication and/or transformation or transfection into host cells and organisms. Examples include plasmids, phages, autonomously replicating sequences (ARS), centromeres, and other sequences which are able to replicate or be replicated in vitro or in a host cell, or to convey a desired nucleic acid segment to a desired location within a host cell. A Vector can have one or more restriction endonuclease recognition sites at which the sequences can be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a nucleic acid fragment can be spliced in order to bring about its replication and cloning. Vectors can further provide primer sites, *e.g.*, for PCR, transcriptional and/or translational initiation and/or regulation sites, recombinational signals, replicons, Selectable markers, *etc.* Clearly, methods of inserting a desired nucleic acid fragment which do not require the use of homologous recombination, transpositions or restriction enzymes (such as, but not limited to, UDG cloning of PCR fragments (US 5,334,575, entirely incorporated herein by refer-

ence), TA Cloning™ brand PCR cloning (Invitrogen Corp., Carlsbad, Calif.), and the like) can also be applied to clone a fragment into a cloning vector to be used according to the present invention. The cloning vector can further contain one or more selectable markers suitable for use in the identification of cells transformed with the cloning vector.

The term "primer" refers to a single stranded or double stranded oligonucleotide that is extended by covalent bonding of nucleotide monomers during amplification or polymerization of a nucleic acid molecule (*e.g.* a DNA molecule). In a preferred aspect, the primer comprises one or more recombination sites or portions of such recombination sites. Portions of recombination sties comprise at least 2 bases, at least 5 bases, at least 10 bases or at least 20 bases of the recombination sites of interest. When using portions of recombination sites, the missing portion of the recombination site may be provided by the newly synthesized nucleic acid molecule. Such recombination sites may be located within and/or at one or both termini of the primer. Preferably, additional sequences are added to the primer adjacent to the recombination site(s) to enhance or improve recombination and/or to stabilize the recombination site during recombination. Such stabilization sequences may be any sequences (preferably G/C rich sequences) of any length. Preferably, such sequences range in size from 1 to about 1,000 bases, 1 to about 500 bases, and 1 to about 100 bases, 1 to about 60 bases, 1 to about 25, 1 to about 10, 2 to about 10 and preferably about 4 bases. Preferably, such sequences are greater than 1 base in length and preferably greater than 2 bases in length.

The term "template" refers to double stranded or single stranded nucleic acid molecules which are to be amplified, synthesized or sequenced. In the case of double stranded molecules, denaturation of its strands to form a first and a second strand is preferably performed before these molecules will be amplified, synthesized or sequenced, or the double stranded molecule may be used directly as a template. For single stranded templates, a primer complementary to a portion of the template is hybridized under appropriate conditions and one or more polypeptides having polymerase activity (*e.g.* DNA polymerases and/or reverse transcriptases) may then synthesize a nucleic acid molecule complementary to all or a portion of said template. Alternatively, for double stranded templates, one or more promoters may be used in combination with one or more polymerases to make nucleic acid molecules complementary to all or a portion of the template. The newly synthesized molecules, according to the invention, may be equal or shorter in length than the original template. Additionally, a population of nucleic acid templates may be used during synthesis or amplification to produce a population of nucleic acid molecules typically representative of the original template population.

The term "adapter" refers to an oligonucleotide or nucleic acid fragment or segment (preferably DNA) which comprises one or more recognition sites (*e.g.*, recombination sites or recognition sites for restriction endonucleases) which can be added to a circular or linear DNA molecule as well as other nucleic acid molecules described herein. Such adapters may be added at any location within a circular or linear molecule, although the adapters are preferably added at or near one or both termini of a linear molecule. Preferably, adapters are positioned to be located on both sides (flanking) a particularly nucleic acid molecule of interest. The synthesis of adapters (*e.g.*, by oligonucleotide synthesis, annealing procedures, and or PCR) is a standard technique

well known to the person skilled in the art. In accordance with the invention, adapters may be added to nucleic acid molecules of interest by standard recombinant techniques (*e.g.* restriction digest and ligation). For example, adapters may be added to a circular molecule by first digesting the molecule with an appropriate restriction enzyme, adding the adapter at the cleavage site and reforming the circular molecule which contains the adapter(s) at the site of cleavage. Alternatively, adapters may be ligated directly to one or more and preferably both termini of a linear molecule thereby resulting in linear molecule(s) having adapters at one or both termini. In one aspect of the invention, adapters may be added to a population of linear molecules, (*e.g.* a cDNA library or genomic DNA which has been cleaved or digested) to form a population of linear molecules containing adapters at one and preferably both termini of all or substantial portion of said population.

Other terms used in the fields of recombinant DNA technology and molecular and cell biology as used herein will be generally understood by one of ordinary skill in the applicable arts.

**DETAILED DESCRIPTION OF THE INVENTION**

A first subject matter of the invention relates to a method for identification and isolation of transcription termination sequences for comprising the steps of:

i) providing a screening construct or screening vector comprising
   a) a promoter sequence, and
   b) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences, and
   c) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic,

   wherein insertion of an efficient transcription terminator into said insertion site changes expression of said additional sequences by said promoter sequence in comparison to no insertion, and

ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and

iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector, and

iv) introducing said screening construct or screening vector with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and

v) identifying and/or selecting screening construct or screening vector with a changed readily detectable characteristic in comparison to no insertion, and

vi) isolating the inserted DNA sequences from said identified and/or selected screening construct or screening vector for use as transcription termination sequences and – optionally – determining their sequence.

By the method of the invention new transcription terminator sequences can be readily identified. The method can be based either on an *in vitro* or *in vivo* screening system. The screening method of the invention allows for selection of DNA sequences, screening constructs or screening vectors, and/or cells or organism (preferably plant cells and plant organisms) containing efficient transcription terminator sequences.

Previously terminators had to be evaluated sequence-by-sequence. Testing of termination efficiency (tightness) was laborious. The methods of the present invention are time-efficient and very sensitive so that only very tight terminators will be identified. Tight terminators identified by this screening system will be used for the expression cassettes, which will reduce read through between cassettes and increase stability of the transgene expression. Discovery of various terminators of interest will provide opportunity to understand better termination of transcription *in planta*.

The term "readily detectable characteristic" as used herein is to be understood in the broad sense and may include any change of a characteristic, preferably a phenotypic characteristic. "Change" in this context may include increasing or decreasing said characteristic. In consequence, expression of said additional sequences under control of said promoter may cause increasing or decreasing a phenotypic characteristic. For example expression may cause a herbicide resistance (increased resistance) or may cause a toxic effect by expression of *e.g.*, a toxic gene (decreased viability). Since depending on the localization of the insertion site in relation to said additional sequences (as described below in detail) an efficient transcription terminator may result in increased (preferably initiated) or decreased (preferably silenced) expression of said additional sequences both type of changes can be advantageously employed.

**1. Localization of the Insertion Site**

The insertion site may have various localizations with respect to the additional sequences which bring about the readily detectable characteristic:

**1.1 Variation A:**

For example the insertion site may be localized upstream (*i.e.* in 5' direction) of the additional sequences so that the insertion site is between the promoter sequences and said additional sequences (hereinafter "Variation A"). In case an efficient transcription terminator sequences in inserted into said insertion site transcription, transcription will stop before said additional sequences and no read-through transcription into this additional sequences will occur. In this case an efficient transcription terminator will result in decreased or preferably completely suppressed expression of the additional sequences. Depending whether presence or absence of transcription (*i.e.* expression) of said additional sequences brings about said readily detectable characteristic (which both is possible depending on the type of additional sequence employed) said readily detectable characteristic diminishes or is expressed. In any case an efficient transcription terminator will cause a changed readily detectable characteristic, which may be suppressed or increased in comparison to a scenario where no sequence is inserted into the insertion site.

Thus in a preferred embodiment of the invention the method for identification and isolation of transcription termination sequences comprises the steps of:

i) providing a screening construct or screening vector comprising in 5' to 3' direction
   a) a promoter sequence, and
   b) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences, and

c) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic,

wherein insertion of an efficient transcription terminator into said insertion site suppresses expression of said additional sequences by said promoter sequence in comparison to no insertion, and

ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and

iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector, and

iv) introducing said screening construct or screening vector with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and

v) identifying and/or selecting screening constructs or screening vectors with a changed readily detectable characteristic in comparison to no insertion, and

vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

As described below the additional sequences localized downstream of the insertion site may bring about said readily detectable characteristic only by expression (*i.e.* transcription) of an RNA (*e.g.*, in cases where said additional sequences are forming an antisense RNA or dsRNA molecule) or by expression (*i.e.* transcription and translation of a protein). In the latter case it has to be ensured that appropriate translation can occur. This can be ensured by for example avoiding upstream ATG-codons, cloning the sequences in-frame with upstream coding sequences or – preferably – employing IRES sites which may allow efficient translation even in cases where the ATG codon is not close to the 5'-end of the transcript (Vagner 2001; for sequences see *e.g.*, http://ifr31w3.toulouse.inserm.fr/IRESdatabase/).

**1.2 Variation B:**

In another preferred embodiment of the invention the insertion site may also be arranged downstream (*i.e.* in 3'-direction) of the additional sequences (hereinafter "Variation B"). For this variation the DNA sequences to be inserted into the insertion sites for evaluation for their transcription termination capability are preferably inserted in form of an inverted repeat. In case the inserted DNA sequence is an efficient transcription terminator only the first copy (*i.e.* first part) of the inverted repeat will be transcribed and normal expression of the additional sequences will occur. Depending whether presence or absence of transcription (*i.e.* expression) of said additional sequences brings about said readily detectable characteristic (which both is possible depending on the type of additional sequence employed) said readily detectable characteristic diminishes or is expressed. In any case an efficient transcription terminator will cause a changed readily detectable characteristic, which may be suppressed or increased in comparison to a scenario where no sequence is inserted into the insertion site. However, if only a weak transcription terminator or a sequence with no transcription termination capability at all is inserted the entire inverted repeat (*i.e.* both copies of the inserted sequence) will be transcribed causing transcription of a RNA comprising a double-stranded hairpin structure (formed by the RNA transcribed from inverted repeat of the inserted DNA sequences). This RNA by means of double-stranded RNA interference (dsRNAi) will

cause gene silencing of its own expression (self-suppression or self-silencing) resulting in gene silencing of the expression cassette comprising said additional sequences. In this case an efficient transcription terminator will result in increased expression of the additional sequences and the detectable characteristic will change in just the other direction as in case of an efficient transcription terminator. Thus in a preferred embodiment of the invention the method for identification and isolation of transcription termination sequences comprises the steps of:

i) providing a screening construct or screening vector comprising in 5' to 3' direction
   a) a promoter sequence, and
   b) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic, and
   c) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences,
ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and
iii) inserting at least two copies of a specific DNA sequence of said DNA sequences in form of an inverted repeat into said insertion site of said screening construct or screening vector, wherein insertion of an inverted repeat of an efficient transcription terminator into said insertion site allows expression of said additional sequences by said promoter sequence in comparison to no insertion, and
iv) introducing said screening constructs or screening vectors with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and
v) identifying and/or selecting screening constructs or screening vectors with said readily detectable characteristic in comparison to no insertion, and
vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

In a preferred embodiment of this variation two different promoters and two different additional sequences are employed. These two expression cassettes are arranged in a "tail-to-tail" orientation so that transcription initiated from said promoters in running against each other. Preferably the insertion site for the inverted repeat in between the two end (tails) of the two expression cassettes. Insertion of an inverted repeat of weak transcription terminator will result of gene silencing of both additional sequences, while insertion of an efficient transcription terminator results in expression of both additional sequences. In consequence a double-check of the transcription termination efficiency becomes feasible. Preferably one of the sequences is selected from the group of negative selection marker (thus an efficient transcription terminator will result in for example a herbicide or antibiotic resistance). The other additional sequence may be selected from the group of reporter genes (for example GFP or GUS; thus an efficient transcription terminator will result in an easily detectable color).

**1.3 Variation C:**

In a third preferred embodiment of the invention the insertion site may also be arranged within the additional sequences (for example and preferably embedded into an intron, which is located in said additional sequences) (hereinafter "Variation C"). The full-length transcript of said additional sequence is only made, if the sequence inserted into said insertion site does not control tight transcription termination and expression of said additional sequences will cause a change of said readily detectable characteristic. In case of an efficient transcription terminator inserted into said insertion site no full-length transcript will be produced. Thus in a preferred embodiment of the invention the method for identification and isolation of transcription termination sequences comprises the steps of:

i) providing a screening construct or screening vector comprising in 5' to 3' direction
   a) a promoter sequence, and
   b) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic, and embedded into said additional sequences one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences,

   wherein insertion of an efficient transcription terminator into said insertion site suppresses full-length transcription of said additional sequences by said promoter sequence in comparison to no insertion, and
ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and
iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector, and
iv) introducing said screening constructs or screening vectors with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and
v) identifying and/or selecting screening constructs or screening vectors with a changed readily detectable characteristic in comparison to no insertion, and
vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

Localization within the additional sequences may be realized by various ways. In an improved embodiment of the invention, the insertion site for the transcription terminator sequences is localized within an intron comprised in said additional sequences. Efficient transcription termination will lead to incomplete transcription of the intron and the additional sequences, thereby preventing the phenotype caused by said additional sequences to occur. The additional sequence to be expressed may be – for example – a toxic gene (such as diphtheria toxin A) or a reporter gene (such as GFP). Additional examples are given below. In case of efficient transcription termination the phenotype corresponding to said sequences is not established. In case of toxic genes only stably transformed cell lines can be established which comprise an efficient transcription terminator sequence.

In another preferred embodiment the insertion site for the transcription termination sequences is localized between a first 5'-part of the additional sequences, which - for example - encodes for a reporter gene (such as GPF) and a 3'-part, which is preferably a non-protein encoding sequence with no homologous sequences in plants (such as for example part of luciferase gene as used in the examples below or sequences of bacteriophage λ). In addition to this screening construct another expression cassette is employed which is expressing an antisense or –preferably – a double-stranded RNA sequence corresponding to said 3'-part sequence. In case of an efficient transcription terminator, transcription of the additional sequence will stop at the terminator site and the 5'-end sequence will not be translated. In case of an inefficient (leaky) transcription terminator transcription will read-through into said 3'-part sequence thereby establishing a target for the antisense- or double-stranded RNA. This will cause degradation of the entire construct, including the region encoding for the marker sequence, thereby "silencing" the related phenotype (*e.g.*, marker signal).

## 2. THE SCREENING CONSTRUCT OR SCREENING VECTOR OF THE INVENTION

The screening constructs and screening vectors to be employed for the method of the invention may have various forms. In principle, any form is suitable which allows for expression or transcription of an RNA molecule. In consequence the screening construct or screening vector may be for example an RNA or a DNA molecule, it further may be single-stranded or double-stranded, and it may be linear or circular. Any combination of the before mentioned alternatives is included.

Screening constructs can be advantageously employed in scenarios were no replication is required, such as for example the *in vitro* screening system described below. However, preferably, a screening vector is employed. Said screening vector may be a RNA vector (such as for example a RNA virus vector) or – preferably – a DNA vector. More preferably the screening vector is a circular double-stranded DNA plasmid vector.

As essential feature the screening construct or screening vector of the invention comprises

a) a promoter sequence, and

b) additional sequences which causes upon expression under said promoter sequence a readily detectable characteristic.

### 2.1 Promoters for the Invention

The promoter is preferably chosen to be functional in the *in vitro* or *in vivo* system where evaluation of said transcription termination sequences is going to be carried out. Preferably, this system is similar or identical to the system where the transcription termination sequence should function in later expression constructs. For example, if a transcription terminator sequence is desired for plant organisms, a transcription system based on plant cells (either an *in vitro* system such as wheat germ extracts or a *in vivo* system such as a plant cell or a plant) is employed. In such a case the promoter sequences is preferably a sequence which is able to initiate transcription in plants, preferably an endogenous plant promoter or a promoter derived from a plant pathogen (such as a plant virus or *Agrobacterium*). Various promoters are known to the person skilled in the art for the various transcriptions systems or hosts for which the method of the invention can be employed.

As an illustration, promoters (and if necessary other transcriptional and translational regulatory signals) suitable for a mammalian host may be derived from viral sources, such as adenovirus, bovine papilloma virus, simian virus, or the like, in which the regulatory signals are associated with a particular gene that has a high level of expression. Suitable transcriptional and translational regulatory sequences also can be obtained from mammalian genes, such as actin, collagen, myosin, and metallothionein genes. Illustrative eukaryotic promoters include the promoter of the mouse metallothionein I gene (Hamer 1982), the TK promoter of Herpes virus (McKnight 1982), the SV40 early promoter (Benoist 1981), the Rous sarcoma virus promoter (Gorman 1982), the cytomegalovirus promoter (Foecking 1980), and the mouse mammary tumor virus promoter (see, generally, Etcheverry 1996). Alternatively, a prokaryotic promoter, such as the bacteriophage T3 RNA polymerase promoter, can be used to control expression of the gene of interest in mammalian cells if the prokaryotic polymerase is expressed by an eukaryotic promoter (Zhou 1990; Kaufman 1991).

For expression in plants, plant-specific promoters are preferred. The term "plant-specific promoter" is understood as meaning, in principle, any promoter which is capable of governing the expression of genes, in particular foreign genes, in plants or plant parts, plant cells, plant tissues or plant cultures. In this context, expression can be, for example, constitutive, inducible or development-dependent. The following are preferred:

a) Constitutive promoters
"Constitutive" promoters refers to those promoters which ensure expression in a large number of, preferably all, tissues over a substantial period of plant development, preferably at all times during plant development. A plant promoter or promoter originating from a plant virus is especially preferably used. The promoter of the CaMV (cauliflower mosaic virus) 35S transcript (Franck 1980; Odell 1985; Shewmaker 1985; Gardner 1986) or the 19S CaMV promoter (US 5,352,605; WO 84/02913) is especially preferred. Another suitable constitutive promoter is the rice actin promoter (McElroy 1990), Rubisco small subunit (SSU) promoter (US 4,962,028), the legumin B promoter (GenBank Acc.No. X03677), the promoter of the nopaline synthase from *Agrobacterium*, the TR dual promoter, the OCS (octopine synthase) promoter from *Agrobacterium*, the ubiquitin promoter (Holtorf S 1995), the ubiquitin 1 promoter (Christensen 1989, 1992; Bruce et al. 1989), the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (US 5,683,439), the promoters of the vacuolar ATPase subunits, the pEMU promoter (Last 1991); the MAS promoter (Velten 1984) and maize H3 histone promoter (Lepetit 1992; Atanassova 1992), the promoter of the *Arabidopsis thaliana* nitrilase-1 gene (GeneBank Acc. No.: U38846, nucleotides 3862 to 5325 or else 5342) or the promoter of a proline-rich protein from wheat (WO 91/13991), and further promoters of genes whose constitutive expression in plants.

b) Tissue-specific or tissue-preferred promoters
Furthermore preferred are promoters with specificities for seeds, such as, for example, the phaseolin promoter (US 5,504,200; Bustos *et al.* 1989; Murai 1983; Sengupta-Gopalan 1985), the promoter of the 2S albumin gene (Joseffson 1987), the legumine promoter (Shirsat 1989), the USP (unknown seed protein) promoter (Bäumlein 1991a), the napin gene promoter (US 5,608,152; Stalberg 1996), the promoter of the sucrose binding proteins (WO 00/26388) or the legumin B4 promoter (LeB4; Bäumlein (1991b),

the *Arabidopsis* oleosin promoter (WO 98/45461), and the *Brassica* Bce4 promoter (WO 91/13980). Further preferred are a leaf-specific and light-induced promoter such as that from cab or Rubisco (Simpson 1985; Timko 1985); an anther-specific promoter such as that from LAT52 (Twell 1989b); a pollen-specific promoter such as that from Zml3 (Guerrero et al. (1993) Mol Gen Genet 224:161-168); and a microspore-preferred promoter such as that from apg (Twell et al. 1983).

c) Chemically inducible promoters

The expression cassettes may also contain a chemically inducible promoter (review article: Gatz 1997), by means of which the expression of the exogenous gene in the plant can be controlled at a particular point in time. Such promoters such as, for example, the PRP1 promoter (Ward 1993), a salicylic acid-inducible promoter (WO 95/19443), a benzenesulfonamide-inducible promoter (EP 0 388 186), a tetracyclin-inducible promoter (Gatz 1991, 1992), an abscisic acid-inducible promoter (EP-A1 0 335 528) or an ethanol-cyclohexanone-inducible promoter (WO 93/21334) can likewise be used. Also suitable is the promoter of the glutathione-S transferase isoform II gene (GST-II-27), which can be activated by exogenously applied safeners such as, for example, N,N-diallyl-2,2-dichloroacetamide (W0 93/01294) and which is operable in a large number of tissues of both monocots and dicots. Further exemplary inducible promoters that can be utilized in the instant invention include that from the ACE1 system which responds to copper (Mett 1993); or the In2 promoter from maize which responds to benzenesulfonamide herbicide safeners (Hershey 1991; Gatz 1994). A promoter that responds to an inducing agent to which plants do not normally respond can be utilized. An exemplary inducible promoter is the inducible promoter from a steroid hormone gene, the transcriptional activity of which is induced by a glucocorticosteroid hormone (Schena 1991).

Particularly preferred are constitutive promoters. Furthermore, further promoters may be linked operably to the nucleic acid sequence to be expressed, which promoters make possible the expression in further plant tissues or in other organisms, such as, for example, *E. coli* bacteria. Suitable plant promoters are, in principle, all of the above-described promoters.

**2.2 Additional Sequences for the Invention**

The "additional sequence", which causes upon expression under said promoter sequence a readily detectable characteristic, can be selected from a broad variety of sequences. Selection may depend on various factors, for example, whether insertion of an efficient transcription terminator into said insertion site is expected to result in decreased expression (Variation A or C) or increased expression (Variation B) or the additional sequences in its functional form (*i.e.*, which brings about the readily detectable characteristic).

For expected decreased expression (Variation A and C) it is preferred to employ a sequence which encodes for a selectable marker selected from the group consisting of a reporter gene, a counter selection marker, or a toxic gene. In an preferred embodiment of the *in vivo* screening systems of the invention, the expression of a toxic gene as the additional sequences will cause a inhibition of growth, propagation and/or or regeneration of said cells or organisms (*e.g.*, plant cells or plants). In consequence, only cells or organisms will survive if an efficient ("tight") transcription termination sequence is in-

serted in front of said toxic phenotype causing sequence thereby preventing expression of this growth, propagation and/or or regeneration inhibiting sequences. The surviving cells can be isolated and the transcription terminator sequence can be identified and isolated, *e.g.*, by amplification using PCR followed by sequencing.

For expected increased expression (Variation B) it is preferred to employ a sequence which encodes for a selectable marker selected from the group consisting of a reporter gene, a negative selection marker, or a positive selection marker.

The term "selection marker" refers to any nucleic acid or amino acid sequence which is useful to select and separate cells or organism comprising said selection marker from others not comprising it. Selection marker may comprise sequences which

i) allow for separation of cells or organism comprising said marker by conferring a resistance against an otherwise toxic compound (named herein within "negative selection marker"),

ii) allow for separation of cells or organism comprising said marker by conferring a growth advantage to said cells or organism (named herein within "positive selection marker").

Selection marker may further comprise sequences which allow for separation of cells or organism not comprising said marker by conferring a growth disadvantage to cells or organism comprising said marker (named herein within "counter selection marker" or "toxic gene").

Selection markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like. Examples of Selectable markers include but are not limited to:

(a) a DNA segment that encodes a product that provides resistance in a recipient cell or organism against otherwise toxic compounds ("Negative Selection Marker"); (*e.g.*, antibiotics). Negative Selection Markers confer a resistance to a biocidal compound such as a metabolic inhibitor (*e.g.*, 2-deoxyglucose-6-phosphate, WO 98/45456), antibiotics (*e.g.*, kanamycin, G 418, bleomycin or hygromycin) or herbicides (*e.g.*, phosphinothricin or glyphosate). Especially preferred Negative Selection Markers are those which confer resistance to herbicides. Examples which may be mentioned are:

- Phosphinothricin acetyltransferases (PAT; also named Bialophos ®resistance; bar; de Block 1987; EP 0 333 033; US 4,975,374)
- 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) conferring resistance to Glyphosate® (N-(phosphonomethyl)glycine) (Shah 1986)
- Glyphosate® degrading enzymes (Glyphosate® oxidoreductase; gox),
- Dalapon® inactivating dehalogenases (deh)
- sulfonylurea- and imidazolinone-inactivating acetolactate synthases (for example mutated ALS variants with, for example, the S4 and/or Hra mutation
- Bromoxynil® degrading nitrilases (bxn)
- Kanamycin- or. G418- resistance genes (NPTII; NPTI) coding *e.g.*, for neomycin phosphotransferases (Fraley *et al.* 1983)
- 2-Desoxyglucose-6-phosphate phosphatase ($DOG^R$1-Gene product; WO

98/45456; EP 0 807 836) conferring resistance against 2-desoxyglucose (Randez-Gil *et al.* 1995).
- hygromycin phosphotransferase (HPT), which mediates resistance to hygromycin (Vanden Elzen et al. 1985).
- dihydrofolate reductase (Eichholtz *et al.* 1987)

Additional negative selectable marker genes of bacterial origin that confer resistance to antibiotics include the *aad*A gene, which confers resistance to the antibiotic spectinomycin, gentamycin acetyl transferase, streptomycin phosphotransferase (SPT), aminoglycoside-3-adenyl transferase and the bleomycin resistance determinant (Hayford 1988; Jones 1987; Svab 1990; Hille 1986).

Especially preferred are negative selection markers which confer resistance against the toxic effects imposed by D-amino acids like *e.g.*, D-alanine and D-serine (WO 03/060133; Erikson 2004). Especially preferred as negative selection marker in this contest are the daol gene (EC: 1.4. 3.3: GenBank Acc.-No.: U60066) from the yeast *Rhodotorula gracilis* (*Rhodosporidium toruloides*) and the *E. coli* gene *dsdA* (D-serine dehydratase (D-serine deaminase) [EC: 4.3. 1.18; GenBank Acc.-No.: J01603).

Selection Marker suitable in prokaryotic or non-plant eukaryotic systems can also be based on the Selection Markers described above for plants (beside that expression cassettes are based on other host-specific promoters). For mammal cells preferred are resistance against neomycin (G418), hygromycin, Bleomycin, Zeocin Gatignol 1987; Drocourt 1990), puromycin (see, for example, Kaufman 1990a, 1990b). Corresponding selectable marker genes are known in the art (see, for example, Srivastava 1991; Romanos 1995; Markie 1996; Pfeifer 1997; Tucker 1997; Hashida-Okado 1998). For prokaryotes preferred are resistances against Ampicillin, Kanamycin, Spectinomycin, or Tetracyclin. Selectable marker genes can be cloned or synthesized using published nucleotide sequences, or marker genes can be obtained commercially.

(b) a DNA segment that encodes a product that is toxic in a recipient cell or organism ("Counter Selection Marker"). A counter selection marker is especially suitable to select organisms with defined deletions originally comprising said marker (Koprek 1999). Examples for negative selection marker comprise thymidin kinases (TK), cytosine deaminases (Gleave 1999; Perera 1993; Stougaard 1993), cytochrome P450 proteins (Koprek 1999), haloalkan dehalogenases (Naested 1999), *iaaH* gene products (Sundaresan 1995), cytosine deaminase *codA* (Schlaman & Hooykaas 1997), or *tms2* gene products (Fedoroff & Smith 1993). In general the term "counter selection marker" within the scope of this invention is to be understood in the broad sense including all proteins which either

i) cause a toxic effect *per se* on the cell or organism (*e.g.*, a plant cell), or

ii) convert a non-toxic compound X into a toxic compound Y.

The term "non toxic compound X" as used herein in intended to mean compounds which in comparison to its conversion product Y - under otherwise identical conditions (*i.e.* conditions which are identical beside the difference in compound X and Y)– demonstrate a reduced, preferably an absent, biological activity, preferably toxicity. Preferably the toxicity of compound Y is at least two-times, preferably at

least five-times, more preferably at least ten-times, most preferably at least one hundred-times the toxicity of the corresponding compound X. Conversion of X to Y can occur by various mechanism including but not limited to hydrolysis, deamination, saporation, dephosphorylation, phosphorylation, oxidation or an other way of activation, metabolization, or conversion. Compound X can for example be an inactive precursor of a plant growth regulator or a herbicide. "Toxicity" or "toxic effect" as used herein means a measurable, negative effect on the physiology of a cell or an organism and may comprise symptoms including but not limited to decreased or impaired growth, decreased or impaired photosynthesis, decreased or impaired cell division, decreased or impaired regeneration or proliferation *etc*.

The counter selection marker may be an endogenous gene or a heterologous gene or transgene from another organism. The following counter selection marker are given by way of example:

1. Cytosine deaminases (*CodA* or CDase), wherein compounds like *e.g.*, 5-fluorocytosine (5-FC) is employed as non-toxic compound X. Cytosine deaminases catalyze deamination of cytosine to uracil (Kilstrup 1989; Anderson 1989). 5-FC is concerted to the toxic metabolite ("Y") 5-fluorouracil (5-FU) (Polak 1975). 5-FC is of low toxicity Toxizität (Bennett 1990). In contrast, 5-FU exhibits a strong cytotoxic effect inhibiting RNA- and DNA-synthesis (Calabrisi 1990; Damon 1989).

   Cells of plants and higher mammals do not exhibit a significant CDase-activity and are unable to deaminate 5-FC (Polak 1976; Koechlin 1966). In the context of the present invention, a CDase is introduced as a transgene into the target cell. Introduction can be done prior the screening (*e.g.*, generating a stably transformed cell line or organism). Such cells or organism can then be used as master cell lines or master organism.

   Corresponding CDase sequences, transgenic organisms (including plants) comprising said sequences, and negative selection schemes based on *e.g.*, treatment of these cells or organisms with 5-FC (as non-toxic substance X) are known in the art (WO 93/01281; US 5,358,866; Gleave 1999; Perera 1993; Stougaard 1993; EP-A1 595 837; Mullen 1992; Kobayashi 1995; Schlaman 1997; Xiaohui Wang 2001; Koprek 1999; Gallego 1999; Salomon 1998; Thykjaer 1997; Serino 1997; Risseeuw 1997; Blanc 1996; Corneille 2001). Cytosindeaminases and genes encoding the same can be isolated from various organisms, preferably microorganism, like for example *Cryptococcus neoformans, Candida albicans, Torulopsis glabrata, Sporothrix schenckii, Aspergillus, Cladosporium*, and *Phialophora* (Bennett 1990) and from bacteria like *e.g.*, *E.coli* and *Salmonella typhimurium* (Andersen 1989). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred are the sequences as described by GenBank Acc.-No: S56903, and the modified sequences described in EP-A1 595 873, which were modified to enable expression in eukaryotes.

2. Cytochrome P-450 enzymes, especially the bacterial cytochrome P-450 SU1 gene product (CYP105A1) from *Streptomyces griseolus* (strain ATCC 11796), wherein substances like the sulfonylurea pro-herbicide R7402 (2-methylethyl-2-3-dihydro-N-[(4,6-dimethoxypyrimidine-2-yl)aminocarbonyl]-1,2-benzoisothiazol-7-sulfonamid-1,1-dioxide) as the non-toxic substance X are employed. Corresponding sequences are negative selection schemes employing *e.g.*, R7402 are described in the art (O'Keefe 1994; Tissier 1999; Koprek 1999; O'Keefe 1991). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No: M32238.

3. Indoleacetic acid hydrolases like *e.g.*, the tms2 gene product from *Agrobacterium tumefaciens*, wherein substances like auxinamide compounds or naphthalacetamide (NAM) are employed as non-toxic compound X (NAM being converted to naphthyacetic acid, a phytotoxic compound). Corresponding sequences and the realization of negative selection schemes (employing NAM as non-toxic compound X) are described in the art (Fedoroff 1993; Upadhyaya 2000; Depicker 1988; Karlin-Neumannn 1991; Sundaresan 1995; Cecchini 1998; Zubko 2000). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No: NC_003308 (Protein_id="NP_536128.1), AE009419, AB016260 (Protein_id="BAA87807.1) and NC002147.

4. Haloalkane dehalogenases (dhlA gene product) *e.g.*, from *Xanthobacter autotropicus* GJ10. This dehalogenase hydrolizes dihaloalkanes like 1,2-dichloroethane (DCE) to halogenated alcohols and inorganic halides (Naested 1999; Janssen 1994; Janssen 1989). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No: M26950.

5. Thymidine kinases (TK), especially virale TKs from virus like Herpes Simplex virus, SV40, Cytomegalovirus, *Varicella zoster* virus, especially preferred is TK from Type 1 Herpes Simplex virus (TK HSV-1), wherein substances like e.g., acyclovir, ganciclovir or 1,2-deoxy-2-fluoro-β-D-arabinofuranosil-5-iodouracile (FIAU) are employed as non-toxic compound X. Corresponding compounds are realization of negative selection schemes (*e.g.*, employing acyclovir, ganciclovir or FIAU) are known in the art (Czako 1994; Wigler 1977; McKnight 1980; McKnight 1980; Preston 1981; Wagner 1981; St. Clair 1987). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No: J02224, V00470, and V00467.

6. Guanine phosphoribosyl transferases, hypoxanthine phosphoribosyl transferases or Xanthin guanine phosphoribosyl transferases, wherein compounds like 6-thioxanthin or allopurinol are employed as non-toxic substance X. Preferred is the guanine phosphoribosyl transferase (gpt) from *e.g. E. Coli* (Besnard 1987; Mzoz 1993; Ono 1997), hypoxanthin phosphoribosyl transferases (HPRT; Jolly 1983; Fenwick 1985), xanthin guanine phosphoribosyl

transferases (*e.g.*, from *Toxoplasma gondii*; Knoll 1998; Donald 1996). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No.: U10247 (*Toxoplasma gondii* HXGPRT), M13422 (*E. coli* gpt) and X00221 (*E. coli* gpt).

7. Purine nucleoside phosphorylases (PNP; *DeoD* gene product) *e.g.*, from *E. coli*, wherein compounds like for example 6-methylpurine deoxyribonucleoside are employed as non-toxic compound X. Suitable compounds and methods for carrying out counter-selection schemes (*e.g.*, employing 6-methylpurine deoxyribonucleoside as non-toxic compound X) are well known to the person skilled in the art (Sorscher 1994). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No.: M60917.

8. Phosphonate monoesterhydrolases, which are suitable to convert physiologically inactive ester derivatives of *e.g.*, the herbicide Glyphosate (*e.g.*, glyceryl-glyphosate) to the active form of the herbicide. Suitable compounds and methods for carrying out counter-selection schemes (*e.g.*, employing glyceryl-glyphosate as non-toxic compound X) are well known to the person skilled in the art (US 5,254,801; Dotson 1996; Dotson 1996). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No.: U44852.

9. Aux-1 and – preferably - Aux-2 gene products *e.g.* aus derived from the Ti-plasmids of *Agrobacterium* strains (Beclin 1993; Gaudin 1995). The activity of both enzymes causes production of indole acetamide (IAA) in the plant cell. Aux-1 is encoding an indole acetamide synthase (IAMS) converting tryptophan to indole acetamide (VanOnckelen 1986). Aux-2 is encoding indole acetamide hydrolase (IAMH) converting indole acetamide (a compound without phytohormon activity) to the active auxin indole acetic acid (Inze 1984; Tomashow 1984; Schroder 1984). IAMH is furthermore capable to convert various indole amide-type substrates such as naphthyl acetamide, which is converted into the plant growth regulator naphthyl acetic acid (NAA). Use of IAMH as counter selection marker is for example disclosed in US 5,180,873. Corresponding enzymes are also described for *A. rhizogenes, A. vitis* (Canaday 1992) and Pseudomonas savastanoi (Yamada 1985). The use as counter selection marker for selectively killing certain plant tissues (*e.g.*, pollen; US 5,426,041) or transgenic plants (US 5,180,873) is described. Compounds and methods for counter selections (*e.g.* by employing naphthyl acetamide) are known to the person skilled in the art (see above). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No.: M61151, AF039169 und AB025110.

10. Adenine phosphoribosyl transferases (APRT), wherein compounds such as 4-aminopyrazolo pyrimidine are employed as non-toxic compound X. Suitable compounds and methods for carrying out counter-selection schemes are well

known to the person skilled in the art (Wigler 1979; Taylor 1985). All these references, the sequences and methods described therein are explicitly incorporated by reference.

11. Methoxinin dehydrogenases, wherein compounds such as 2-amino-4-methoxy-butanicacid (Methoxinin) are employed as non-toxic compound X, which is converted into the toxic compound methoxyvinylglycine (Margraff 1980).

12. Rhizobitoxine synthases, wherein compound such as 2-amino-4-methoxy-butanicacid (Methoxinin) are employed as non-toxic compound X, which is converted into the toxic compound 2-amino-4-[2-amino-3-hydroxypropyl]-trans-3-butanicacid (Rhizobitoxin) (Owens 1973).

13. 5-Methylthioribose (MTR) kinases, wherein compounds such as 5-trifluoromethyl thioribose (MTR-analogue, "subversives substrate") are employed as non-toxic compound X, which is converted into the toxic compound Y carbothionyldifluoride. MTR-kinase is a key enzyme of the methionine salvage pathway. Corresponding enzyme activities are described in plants, bacteria, and protozoa but not in mammals. MTR kinases from various species can be identified according to defined sequence motives (Sekowska 2001; http://www.biomedcentral.com/1471-2180/1/15). Corresponding sequences are methods for counter selection (*e.g.*, employing 5-trifluoromethyl thioribose) are known to the person skilled in the art and readily obtainable from sequence databases (*e.g.*, Sekowska 2001; Cornell 1996). All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No.: AF212863 or AC079674 and other MTK kinase enzymes as described in WO 03/078629 and DE10212892 hereby incorporated by reference.

14. Alcohol dehydrogenases (Adh) especially plant Adh-1 gene products, where preferably compounds such as allylalcohol is employed as non-toxic compound X, which is converted into the toxic compound (Y) acrolein. Suitable, corresponding compounds and methods for carrying out counter-selection schemes (*e.g.*, employing allyl alcohol) are well known to the person skilled in the art (Wisman 1991; Jacobs 1988; Schwartz 1981). Sequences can be readily derived from sequence databases. All these references, the sequences and methods described therein are explicitly incorporated by reference. Especially preferred is the sequence described by GenBank Acc.-No.: X77943, M12196, AF172282, X04049 or AF253472.

15. Furthermore preferred as counter selection marker are "toxic genes" or "toxic sequences" which *per se* exhibit and toxic effect on a cell expressing said genes or sequences. Example may include but are not limited to sequences encoding toxic protein such as diphtheria toxin A, Ribonukleases (RNAse *e.g.*, Barnase), ribosome-inhibiting proteins (RIP; such as ricine), magainins, DNAse, phytotoxins, proteins which are able to evoke a hypersensitive reaction, and proteases. Evoking a hypersensitive response (HR) is possible when a pathogen-derived elicitor protein and a corresponding plant-derived receptor

protein are expressed simultaneously. Couples of such corresponding elicitor/receptor genes and their applicability to evoke a HR in a transgenic plant, are known in the art, *e.g.* for *Cladosporium fulvum avr*-genes and *Lycopersicon esculentum Cf*-genes (WO 91/15585) or for *Psuedomonas syringae* avr-genes and *Arabidopsis thaliana RPM1*-genes (Grant 1995).

Additional toxic sequences are those suppressing essential endogenous genes (such as housekeeping genes). The person skilled in the art is aware of various sequences and methods which can be employed to suppress ("silence") gene expression of endogenous genes. The terms "suppression" or "silencing" in relation to a gene, its gene product, or the activity of said gene product is to be understood in the broad sense comprising various mechanism of impairing or reducing the functionality on various levels of expression. Included are for example a quantitative reduction of transcription and translation up to an essentially complete absence of the transcription and/or translation product (*i.e.* lacking detectability by employing detection methods such as Northern or Western blot analysis, PCR, *etc.*)

Suitable method of gene silencing may include but shall not be limited to gene silencing by

(i) antisense suppression (see above for details),

(ii) sense suppression (co-suppression) (see above for details),

(iii) double-stranded RNA interference (see above for details),

(iv) expression of ribozymes against an endogenous RNA transcript (see above for details),

(v) expression of protein or DNA-binding factors: Expression of an endogenous gene can be "silenced" also by expression of certain DNA or protein binding factors which interfere with expression or activity of the gene of its gene product. For example artificial transcription factors of the zinc finger type can be adapted to any target sequence and can thus be employed for gene silencing (*e.g.*, by being directed against the promoter region of the target gene). Methods for production of such factors are described (Dreier 2001; Dreier 2000; Beerli 2000a, 2000b; Segal 2000; Kang 2000; Beerli 1998; Kim 1997; Klug 1999; Tsai 1998; Mapp 2000; Sharrocks 1997; Zhang 2000). Furthermore factors can be employed which directly inhibit the gene product (by interacting with the resulting protein). Such protein binding factors may for example be aptameres (Famulok 1999), antibodies, antibody fragments, or single chain antibodies. Their generation is described (Owen 1992; Franken 1997; Whitelam 1996).

(vi) Gene silencing mediating viral expression systems: Gene silencing of endogenous genes can also be mediated employing specific viral expressions systems (Amplikon; Angell 1999). These systems and methods (termed "VIGS"; viral induced gene silencing) are mediating expression of sequences resembling the endogenous gene from a viral vector system. By classifying the expression as "viral" the entire expression (including expression of the homologous endogenous gene) is shot down by plant viral defense mechanism. Corresponding methods are described in the art (Ratcliff 2001; Fagard 2000; Anandalakshmi 1998; Ruiz 1998).

Essential endogenous genes suitable as targets for the method of the invention may for example be genes selected from those coding for enzymes that are essential for cell viability. These so called "housekeeping genes" may for example be selected from genes encoding for proteins such as ATP synthase, cytochrome c, pyruvate kinase, aminoacyl transferase, or phosphate, di-, tricarboxylkate and 2-oxo-glutarate translocators. A list of target enzymes is given in Table 1 by way of example but the invention is not limited to the enzymes mentioned in this table. More detailed listings can be assembled from series as Biochemistry of Plants (Eds. Stumpf & Conn, 1988-1991, Vols. 1-16 Academic Press) or Encyclopedia of Plant Physiology (New Series, 1976, Springer-Verlag, Berlin).

Table 1: EXAMPLES OF TARGET ENZYMES

| Enzvme |
|---|
| ATP synthase (mitochondrion) |
| adenine nucleotide translocator (mitochondrion) |
| phosphate translocator (mitochondrion) |
| tricarboxylate translocator (mitochondrion) |
| dicarboxylate translocator (mitochondrion) |
| 2-oxo-glutarate translocator (mitochondrion) |
| cytochrome C (mitochondrion) |
| pyruvate kinase |
| glyceraldehyde-3P-dehydrogenase |
| NADPH-cytochrome P450 reductase |
| fatty acid synthase complex |
| glycerol-3P-acyltransferase |
| hydroxymethyl-glutaryl CoA reductase |
| aminoacyl transferase |
| transcription factors |
| elongation factors |
| phytoen desaturase |
| nitrate reductase |
| p-hydroxyphenylpyruvate dioxygenase (HPPD) |
| transketolase<br>(preferably enzymes described and claimed in EP-A1 723 017) |
| ferredoxin oxidoreductase<br>(preferably enzymes described and claimed in EP-A1 1 333 098) |
| S-adenosylmethionin:Mg-protoporphyrin-IX-O-methyltransferase<br>(preferably enzymes described and claimed in EP-A1 1 198 578) |
| dihydrorotase (EC 3.5.2.3)<br>(preferably enzymes described and claimed in EP-A1 1 210 437) |
| phosphoribosyl pyrophosphate synthase<br>(preferably enzymes described and claimed in EP-A1 1 294 925) |
| aspartate carbamyl transferase<br>(preferably enzymes described and claimed in EP-A1 1 259 623) |
| dehydrochinate dehydratase / shikimate dehydrogenase<br>(preferably enzymes described and claimed in EP-A1 1 315 808 |

As housekeeping genes are in general highly conserved, heterologous probes from other (plant) species can be used to isolate the corresponding gene from the species that is to be made resistant. Such gene isolations are well within reach of those skilled in the art and, in view of the present teaching require no undue experimentation.

(c) a DNA segment that encodes a product conferring to the recipient cell or organism an advantage by increased or improved regeneration, growth, propagation, multiplication ("Positive Selection Marker"). Genes like isopentenyl transferase from *Agrobacterium tumefaciens* (strain:PO22; Genbank Acc.-No.: AB025109) may – as a key enzyme of the cytokinin biosynthesis – facilitate regeneration of transformed plants (*e.g.*, by selection on cytokinin-free medium). Corresponding selection methods are described (Ebinuma *et al.* 2000a, 2000b). Additional Positive Selection Markers, which confer a growth advantage to a transformed plant in comparison with a non-transformed one, are described *e.g.,* in EP-A 0 601 092. Growth stimulation selection markers may include (but shall not be limited to) β-glucuronidase (in combination with *e.g.*, a cytokinin glucuronide), mannose-6-phosphate isomerase (in combination with mannose), UDP-galactose-4-epimerase (in combination with *e.g.*, galactose), wherein mannose-6-phosphate isomerase in combination with mannose is especially preferred.

(d) a DNA segment that encodes a product that can be readily identified ("reporter genes" or "reporter proteins" or "reporter molecules"; *e.g.*, phenotypic markers such as β-galactosidase, green fluorescent protein (GFP), and cell surface proteins). The term "reporter gene", "reporter protein", or "reporter molecule" is intended to mean any readily quantifiable protein (or the sequence encoding therefore), which via – for example - color or enzyme activity, makes possible an assessment of presence of said protein or expression of said reporter gene. Reporter genes encode readily quantifiable proteins and, via their color or enzyme activity, make possible an assessment of the transformation efficacy, the site of expression or the time of expression. Very especially preferred in this context are genes encoding reporter proteins (Schenborn 1999) such as the green fluorescent protein (GFP) (Sheen 1995; Haseloff 1997; Reichel 1996; Tian 1997; WO 97/41228; Chui 1996; Leffel 1997), the NAN reporter gene (Kavanagh 2002; WO 03/052104), chloramphenicol transferase, a luciferase (Ow 1986; Millar 1992), the aequorin gene (Prasher 1985), β–galactosidase, R locus gene (encoding a protein which regulates the production of anthocyanin pigments (red coloring) in plant tissue and thus makes possible the direct analysis of the promoter activity without addition of further auxiliary substances or chromogenic substrates (Dellaporta 1988; Ludwig 1990), with β–glucuronidase (GUS) being very especially preferred (Jefferson 1987b; 1987a). β-glucuronidase (GUS) expression is detected by a blue color on incubation of the tissue with 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid, bacterial luciferase (LUX) expression is detected by light emission; firefly luciferase (LUC) expression is detected by light emission after incubation with luciferin; and galactosidase expression is detected by a bright blue color after the tissue is stained with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside. Reporter genes may also be used as scorable markers as alternatives to antibiotic resistance markers. Such markers are used to detect the presence or to measure the level of

expression of the transferred gene. The use of scorable markers in plants to identify or tag genetically modified cells works well only when efficiency of modification of the cell is high.

(e) a DNA segment that encodes a product that inhibits a cell function in a recipient cell;

(f) a DNA segment that inhibits the activity of any of the DNA segments of (a)-(e) above;

(g) a DNA segment that binds a product that modifies a substrate (*e.g.* restriction endonucleases);

(h) a DNA segment that encodes a specific nucleotide recognition sequence which can be recognized by a protein, an RNA, a DNA or a chemical,

(i) a DNA segment that, when deleted or absent, directly or indirectly confers resistance or sensitivity to cell killing by particular compounds within a recipient cell;

(j) a DNA segment that encodes a product that suppresses the activity of a gene product in a recipient cell;

(k) a DNA segment that encodes a product that is otherwise lacking in a recipient cell (*e.g*, tRNA genes, auxotrophic markers), and;

(l) a DNA segment that can be used to isolate or identify a desired molecule (*e.g.*, specific protein binding sites).

In a preferred embodiment of the invention, in cases where an efficient transcription terminator would lead to a decreased expression of the additional sequences (especially Method A, where the transcription terminator is inserted between said additional sequences and the promoter), said additional sequences may by an inverted repeat of a known transcription terminator sequence (such as for example the *nos* terminator) which is localized in a way that the second copy (the copy more downstream from the promoter sequence) is in its "normal" orientation (in which it is constituting a functional transcription terminator). It has to be noted, that such decreased expression of an inverted repeat transcription terminator leads to increased expression (or better not silenced expression) of sequences localized upstream of the insertion site. In case these sequences are encoding for example a marker, an increased resistance or signal can be observed.

Preferably, this transcription terminator (hereinafter "the second transcription terminator") is different from the sequence to be assessed for its transcription termination efficiency. In this case, it is preferred that further sequences are employed which are preferably localized between the promoter and the insertion site and are encoding *e.g.*, for a selection marker or a reporter gene. In such a scenario, an efficient transcription terminator sequence would stop transcription and would not cause transcription of the inverted repeat of said second transcription terminator. In consequence normal expression of the sequences between promoter and insertion site would occur (leading to expression of the selection marker or the reporter gene). In cases, where the sequence inserted into the insertion site is not an effective transcription terminator, transcription will read-through into the inverted repeats of said second transcription terminator. Such a construct would cause its own gene silencing be dsRNAi. In consequence no expression of the sequences between promoter and insertion site would occur (silencing the expression of the selection marker or the reporter gene). Self-affecting gene silencing

based on an inverted repeat sequence of an transcription terminator (*e.g.*, NOS terminator) are described (Brummell 2003).

**2.3 Other Elements of the Screening Construct or Screening Vector**

The screening construct or screening vector may comprise further elements (*e.g.*, genetic control sequences) in addition to a promoter and the additional sequences. The term "genetic control sequences" is to be understood in the broad sense and refers to all those sequences which have an effect on the materialization or the function of the screening construct or screening vector according to the invention. For example, genetic control sequences modify the transcription and translation in prokaryotic or eukaryotic organisms. Genetic control sequences furthermore also encompass the 5'-untranslated regions, introns or noncoding 3'-region of genes, such as, for example, the actin-1 intron, or the Adh1-S introns 1, 2 and 6 (general reference: The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994)). It has been demonstrated that they may play a significant role in the regulation of gene expression. Thus, it has been demonstrated that 5'-untranslated sequences can enhance the transient expression of heterologous genes. Examples of translation enhancers which may be mentioned are the tobacco mosaic virus 5'-leader sequence (Gallie, 1987) and the like. Furthermore, they may promote tissue specificity (Rouster, 1998).

The screening construct or screening vector may advantageously comprise one or more enhancer sequences, linked operably to the promoter, which make possible an increased recombinant expression of the nucleic acid sequence. Additional advantageous sequences, such as further regulatory elements or additional transcription terminator sequences, may also be inserted at the 3'-end of the nucleic acid sequences to be expressed recombinantly.

In some embodiments of the invention (for example Variation A or C) or screening construct or screening vector can also include a known transcription termination sequence (preferably after the additional sequence), and optionally, a polyadenylation signal sequence. Polyadenylation signals which are suitable as control sequences are plant polyadenylation signals, preferably those which essentially correspond to T-DNA polyadenylation signals from *Agrobacterium tumefaciens*, in particular the OCS (octopine synthase) terminator and the NOS (nopaline synthase) terminator. An expression vector does not necessarily need to contain transcription termination and polyadenylation signal sequences, because these elements can be provided by the cloned gene or gene fragment.

The screening construct or screening vector of the invention may comprise further functional elements. The term functional element is to be understood in the broad sense and refers to all those elements which have an effect on the generation, amplification or function of the screening construct or screening vector according to the invention. Functional elements may include for example (but shall not be limited to) selectable marker genes (including negative, positive, and counter selection marker, see above for details), reporter genes, and

1) Origins of replication, which ensure amplification of the expression cassettes or vectors according to the invention in, for example, *E. coli*. Examples which may be mentioned are ORI (origin of DNA replication), the pBR322 ori or the P15A ori (Maniatis

1989). Additional examples for replication systems functional in *E. coli*, are ColE1, pSC101, pACYC184, or the like. In addition to or in place of the E. coli replication system, a broad host range replication system may be employed, such as the replication systems of the P-1 Incompatibility plasmids; *e.g.*, pRK290. These plasmids are particularly effective with armed and disarmed Ti-plasmids for transfer of T-DNA to the plant species host. An expression vector can also include a SV40 origin. This element can be used for episomal replication and rescue in cell lines expressing SV40 large T antigen.

2) Elements which are necessary for *Agrobacterium*-mediated plant transformation, such as, for example, the right and/or – optionally - left border of the T-DNA or the *vir* region.

3) Cloning Sites: The cloning site can preferably be a multicloning site. Any multicloning site can be used, and many are commercially available.

4) S/MAR (scaffold/matrix attachment regions). Matrix attachment regions (MARs) are operationally defined as DNA elements that bind specifically to the nuclear matrix (nuclear scaffold proteins) *in vitro* and are proposed to mediate the attachment of chromatin to the nuclear scaffold *in vivo*. It is possible, that they also mediate binding of chromatin to the nuclear matrix *in vivo* and alter the topology of the genome in interphase nuclei. When MARs are positioned on either side of a transgene their presence usually results in higher and more stable expression in transgenic organisms (especially plants) or cell lines, most likely by minimizing gene silencing (for reveiw: Allen 2000). Various S/MARS sequences and there effect on gene expression are described (Sidorenko 2003; Allen 1996; Villemure 2001; Mlynarova 2002). S/MAR elements may be preferably employed to reduce unintended gene silencing (Mlynarova 2003). An example for a S/MAR being the chicken lysozyme A element (Stief 1989).

5) Sequences which further modify transcription, translation, and/or transport of an expressed protein. For example the expressed protein may be a chimeric protein comprising a secretory signal sequence. The secretory signal sequence is operably linked to a gene of interest such that the two sequences are joined in the correct reading frame and positioned to direct the newly synthesized polypeptide of interest into the secretory pathway of the host cell. Secretory signal sequences are commonly positioned 5' to the nucleotide sequence encoding the amino acid sequence of interest, although certain secretory signal sequences may be positioned elsewhere in the nucleotide sequence of interest (US 5,037,743, US 5,143,830). Expression vectors can also comprise nucleotide sequences that encode a peptide tag to aid the purification of the polypeptide of interest. Peptide tags that are useful for isolating recombinant polypeptides include poly-Histidine tags (which have an affinity for nickel-chelating resin), c-myc tags, calmodulin binding protein (isolated with calmodulin affinity chromatography), substance P, the RYIRS tag (which binds with anti-RYIRS antibodies), the Glu-Glu tag, and the FLAG tag (which binds with anti-FLAG antibodies; see, for example, Luo 1996; Morganti 1996, and Zheng 1997). Nucleic acid molecules encoding such peptide tags are available, for example, from Sigma-Aldrich Corporation (St. Louis, Mo.).

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the composition, methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.

**2.4. Suitable Vectors for the Invention**

As used herein, the terms "vector" and "vehicle" are used interchangeably in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another. The term "screening vector" as used herein refers to a recombinant DNA molecule comprising at least the above defined elements of said promoter and said additional sequences functional for evaluation of the transcription termination efficiency of an inserted sequence.

The methods of the invention are not limited to the vectors disclosed herein. Any vector which is capable of expressing a nucleic acid sequences, and preferably introducing a nucleic acid sequence of interest into a cell (*e.g.*, a plant cell) is contemplated to be within the scope of this invention. Typically, vectors comprise the above defined essential elements of the invention in combination with elements which allow cloning of the vector into a bacterial or phage host. The vector preferably, though not necessarily, contains an origin of replication which is functional in a broad range of prokaryotic hosts. A selectable marker is generally, but not necessarily, included to allow selection of cells bearing the desired vector. Examples of vectors may be plasmids, cosmids, phages, viruses or *Agrobacteria*. More specific examples are given below for the individual transformation technologies.

Preferred are those vectors which make possible a stable integration of the expression construct into the host genome. In the case of injection or electroporation of DNA into cells (*e.g.*, plant cells), the plasmid used need not meet any particular requirements. Simple plasmids such as those of the pUC series can be used. If intact plants are to be regenerated from the transformed cells, it is necessary for an additional selectable marker gene to be present on the plasmid. A variety of possible plasmid vectors are available for the introduction of foreign genes into plants, and these plasmid vectors contain, as a rule, a replication origin for multiplication in *E.coli* and a marker gene for the selection of transformed bacteria. Examples are pBR322, pUC series, M13mp series, pACYC184 and the like.

Preferred vectors for use in the invention include prokaryotic vectors, eukaryotic vectors or vectors which may shuttle between various prokaryotic and/or eukaryotic systems (*e.g.* shuttle vectors). Preferred eukaryotic vectors comprise vectors, which replicate in yeast cells, plant cells, fish cells, eukaryotic cells, mammalian cells, or insect

cells. Preferred prokaryotic vectors comprise vectors which replicate in gram negative and/or gram-positive bacteria, more preferably vectors which replicate in bacteria of the genus *Escherichia, Salmonella, Bacillus, Streptomyces, Agrobacterium, Rhizobium*, or *Pseudomonas*. Most preferred are vectors which replicates in both *E. coli* and *Agrobacterium*. Eukaryotic vectors for use in the invention include vectors which propagate and/or replicate and yeast cells, plant cells, mammalian cells (particularly human cells), fungal cells, insect cells, fish cells and the like. Particular vectors of interest include but are not limited to cloning vectors, sequencing vectors, expression vectors, fusion vectors, two-hybrid vectors, gene therapy vectors, and reverse two-hybrid vectors. Such vectors may be used in prokaryotic and/or eukaryotic systems depending on the particular vector.

In accordance with the invention, any vector may be used to construct a screening vector of the invention. In particular, vectors known in the art and those commercially available (and variants or derivatives thereof) may in accordance with the invention be engineered to include one or more recombination sites for use in the methods of the invention. Such vectors may be obtained from, for example, Invitrogen, Vector Laboratories Inc., Invitrogen, Promega, Novagen, NEB, Clontech, Boehringer Mannheim, Pharmacia, EpiCenter, OriGenes Technologies Inc., Stratagene, PerkinElmer, Pharmingen, Life Technologies, Inc., and Research Genetics. Such vectors may then for example be used for cloning or subcloning nucleic acid molecules of interest. General classes of vectors of particular interest include prokaryotic and/or eukaryotic cloning vectors, expression vectors, fusion vectors, two-hybrid or reverse two-hybrid vectors, shuttle vectors for use in different hosts, mutagenesis vectors, transcription vectors, vectors for receiving large inserts and the like.

Other vectors of interest include viral origin vectors (M13 vectors, bacterial phage λ vectors, adenovirus vectors, and retrovirus vectors), high, low and adjustable copy number vectors, vectors which have compatible replicons for use in combination in a single host (pACYC184 and pBR322) and eukaryotic episomal replication vectors (pCDM8).

Particular vectors of interest include prokaryotic expression vectors such as pcDNA II, pSL301, pSE280, pSE380, pSE420, pTrcHisA, B, and C, pRSET A, B, and C (Invitrogen, Inc.), pGEMEX-1, and pGEMEX-2 (Promega, Inc.), the pET vectors (Novagen, Inc.), pTrc99A, pKK223-3, the pGEX vectors, pEZZ18, pRIT2T, and pMC1871 (Pharmacia, Inc.), pKK233-2 and pKK388-1 (Clontech, Inc.), and pProEx-HT (Life Technologies, Inc.) and variants and derivatives thereof Vector donors can also be made from eukaryotic expression vectors such as pFastBac, pFastBacHT, pFastBacDUAL, pSFV, and pTet-Splice (Life Technologies, Inc.), pEUK-C1, pPUR, pMAM, pMAMneo, pBI101, pBI121, pDR2, pCMVEBNA, and pYACneo (Clontech), pSVK3, pSVL, pMSG, pCH110, and pKK232-8 (Pharmacia, Inc.), p3'SS, pXT1, pSG5, pPbac, pMbac, pMC1neo, and pOG44 (Stratagene, Inc.), and pYES2, pAC360, pBlueBacHis A, B, and C, pVL1392, pBsueBacIII, pCDM8, pcDNA1, pZeoSV, pcDNA3 pREP4, pCEP4, and pEBVHis (Invitrogen, Inc.) and variants or derivatives thereof.

Other vectors of particular interest include pUC18, pUC19, pBlueScript, pSPORT, cosmids, phagemids, YAC's (yeast artificial chromosomes), BAC's (bacterial artificial chromosomes), P1 (*E. coli* phage), pQE70, pQE60, pQE9 (quagan), pBS vectors, PhageScript vectors, BlueScript vectors, pNH8A, pNH16A, pNH18A, pNH46A (Strata-

gene), pcDNA3 (Invitrogen), pGEX, pTrsfus, pTrc99A, pET-5, pET-9, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia), pSPORT1, pSPORT2, pCMVSPORT2.0 and pSV-SPORT1 (Life Technologies, Inc.) and variants or derivatives thereof.

Additional vectors of interest include pTrxFus, pThioHis, pLEX, pTrcHis, pTrcHis2, pRSET, pBlueBacHis2, pcDNA3.1/His, pcDNA3.1(-)/Myc-His, pSecTag, pEBVHis, pPIC9K, pPIC3.5K, pAO815, pPICZ, pPICZα, pGAPZ, pGAPZα, pBlueBac4.5, pBlue-BacHis2, pMelBac, pSinRep5, pSinHis, pIND, pIND (SP1), pVgRXR, pcDNA2.1. pYES2, pZErO1.1, pZErO-2.1, pCR-Blunt, pSE280, pSE380, pSE420, pVL1392, pVL1393, pCDM8, pcDNA1.1, pcDNA1.1/Amp, pcDNA3.1, pcDNA3.1/Zeo, pSe, SV2, pRc/CMV2, pRc/RSV, pREP4, pREP7, pREP8, pREP9, pREP 10, pCEP4, pEBVHis, pCR3.1, pCR2.1, pCR3.1-Uni, and pCRBac from Invitrogen; λExCell, λgt11, pTrc99A, pKK223-3, pGEX-1λT, pGEX-2T, pGEX-2TK, pGEX-4T-1, pGEX-4T-2, pGEX-4T-3, pGEX-3X, pGEX-5X-1, pGEX-5X-2, pGEX-5X-3, pEZZ18, pRIT2T, pMC1871, pSVK3, pSVL, pMSG, pCH110, pKK232-8, pSL1180, pNEO, and pUC4K from Pharmacia; pSCREEN-1b(+), pT7Blue(R), pT7Blue-2, pCITE-4abc(+), pOCUS-2, pTAg, pET-32LIC, pET-30LIC, pBAC-2cp LIC, pBACgus-2cp LIC, pT7Blue-2 LIC, pT7Blue-2, λSCREEN-1, λBlueSTAR, pET-3abcd, pET-7abc, pET9abcd, pET11abcd, pET12abc, pET-14b, pET-15b, pET-16b, pET-17b-pET-17xb, pET-19b, pET-20b(+), pET-21abcd(+), pET-22b(+), pET-23abcd(+), pET-24abcd(+), pET-25b(+), pET-26b(+), pET-27b(+), pET-28abc(+), pET-29abc(+), pET-30abc(+), pET-31b(+), pET-32abc(+), pET-33b(+), pBAC-1, pBACgus-1, pBAC4x-1, pBACgus4x-1, pBAC-3cp, pBACgus-2cp, pBACsurf-1, plg, Signal plg, pYX, Selecta Vecta-Neo, Selecta Vecta-Hyg, and Selecta Vecta-Gpt from Novagen; pLexA, pB42AD, pGBT9, pAS2-1, pGAD424, pACT2, pGAD GL, pGAD GH, pGAD10, pGilda, pEZM3, pEGFP, pEGFP-1, pEGFP-N, pEGFP-C, pEBFP, pGFPuv, pGFP, p6xHis-GFP, pSEAP2-Basic, pSEAP2-Contral, pSEAP2-Promoter, pSEAP2-Enhancer, pβgal-Basic, pβgal-Control, pβgal-Promoter, pβgal-Enhancer, pCMVβ, pTet-Off, pTet-On, pTK-Hyg, pRetro-Off, pRetro-On, pIRES1neo, pIRES1hyg, pLXSN, pLNCX, pLAPSN, pMAMneo, pMAMneo-CAT, pMAMneo-LUC, pPUR, pSV2neo, pYEX4T-1/2/3, pYEX-S1, pBacPAK-His, pBacPAK8/9, pAcUW31, BacPAK6, pTriplEx, λgt10, λgt11, pWE15, and λTriplEx from Clontech; Lambda ZAP II, pBK-CMV, pBK-RSV, pBluescript II KS +/-, pBluescript II SK +/-, pAD-GAL4, pBD-GAL4 Cam, pSurfscript, Lambda FIX II, Lambda DASH, Lambda EMBL3, Lambda EMBL4, SuperCos, pCR-Scrigt Amp, pCR-Script Cam, pCR-Script Direct, pBS +/-, pBC KS +/-, pBC SK +/-, Phagescript, pCAL-n-EK, pCAL-n, pCAL-c, pCAL-kc, pET-3abcd, pET-11'abcd, pSPUTK, pESP-1, pCMVLacI, pOPRSVI/MCS, pOPI3 CAT,pXT1, pSG5, pPbac, pMbac, pMC1neo, pMC1neo Poly A, pOG44, pOG45, pFRTβGAL, pNEOβGAL, pRS403, pRS404, pRS405, pRS406, pRS413, pRS414, pRS415, and pRS416 from Stratagene.

Two-hybrid and reverse two-hybrid vectors of particular interest include pPC86, pDBLeu, pDBTrp, pPC97, p2.5, pGAD1-3, pGAD10, pACt, pACT2, pGADGL, pGADGH, pAS2-1, pGAD424, pGBT8, pGBT9, pGAD-GAL4, pLexA, pBD-GAL4, pHISi, pHISi-1, placZi, pB42AD, pDG202, pJK202, pJG4-5, pNLexA, pYESTrp and variants or derivatives thereof.

Preferred vectors for expression in *E.coli* are pQE70, pQE60 und pQE-9 (QIAGEN, Inc.); pBluescript vectors, phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A

(Stratagene Cloning Systems, Inc.); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia Biotech, Inc.).

Preferred vectors for expression in eukaryotic, animals systems comprise pWLNE0, pSV2CAT, pOG44, pXT1 and pSG (Stratagene Inc.); pSVK3, pBPV, pMSG und pSVL (Pharmacia Biotech, Inc.). Examples for inducible vectors are pTet-tTak, pTet-Splice, pcDNA4/TO, pcDNA4/TO/LacZ, pcDNA6/TR, pcDNA4/TO/Myc-His/LacZ, pcDNA4/TO/Myc-His A, pcDNA4/TO/Myc-His B, pcDNA4/TO/Myc-His C, pVgRXR (Invitrogen, Inc.) or the pMAM-Serie (Clontech, Inc.; GenBank Accession No.: U02443).

Preferred vectors for the expression in yeast comprise for example pYES2, pYD1, pTEFI/Zeo, pYES2/GS, pPICZ, pGAPZ, pGAPZalph, pPIC9, pPIC3.5, PHIL-D2, PHIL-SI, pPIC3SK, pPIC9K, and PA0815 (Invitrogen, Inc.).

Preferred vector for plant transformation are described herein below and preferably comprise vectors for *Agrobacterium*-mediated transformation. *Agrobacterium tumefaciens* and *A. rhizogenes* are plant-pathogenic soil bacteria, which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of the plant (Kado 1991). Vectors of the invention may be based on the *Agrobacterium* Ti- or Ri-plasmid and may thereby utilize a natural system of DNA transfer into the plant genome.

As part of this highly developed parasitism *Agrobacterium* transfers a defined part of its genomic information (the T-DNA; flanked by about 25 bp repeats, named left and right border) into the chromosomal DNA of the plant cell (Zupan 2000). By combined action of the so-called *vir* genes (part of the original Ti-plasmids) said DNA-transfer is mediated. For utilization of this natural system, Ti-plasmids were developed which lack the original tumor inducing genes ("disarmed vectors"). In a further improvement, the so called "binary vector systems", the T-DNA was physically separated from the other functional elements of the Ti-plasmid (*e.g.*, the *vir* genes), by being incorporated into a shuttle vector, which allowed easier handling (EP-A1 0 120 516; US 4,940,838). These binary vectors comprise (beside the disarmed T-DNA with its border sequences), prokaryotic sequences for replication both in *Agrobacterium* and *E. coli*. It is an advantage of *Agrobacterium*-mediated transformation that in general only the DNA flanked by the borders is transferred into the genome and that preferentially only one copy is inserted. Descriptions of *Agrobacterium* vector systems and methods for *Agrobacterium*-mediated gene transfer are known in the art (Miki 1993; Gruber 1993; Moloney 1989). The use of T-DNA for the transformation of plant cells has been studied and described intensively (EP-A1 120 516; Hoekema 1985; Fraley 1985; and An 1985). Various binary vectors are known, some of which are commercially available such as, for example, pBIN19 (Clontech Laboratories, Inc. USA).

Hence, for *Agrobacterium*-mediated transformation the screening construct may be integrated into or the screening vector may consist of specific plasmids, such as shuttle or intermediate vectors, or binary vectors. If a Ti or Ri plasmid is to be used for the transformation, at least the right border, but in most cases the right and left border, of the Ti or Ri plasmid T-DNA is linked to the transgenic expression construct to be introduced in the form of a flanking region. Binary vectors are preferably used. Binary vectors are capable of replication both in *E.coli* and in *Agrobacterium*. They may comprise

a selection marker gene and a linker or polylinker (for insertion of *e.g.* the expression construct to be transferred) flanked by the right and left T-DNA border sequence. They can be transferred directly into *Agrobacterium* (Holsters 1978). The selection marker gene permits the selection of transformed *Agrobacteria* and is, for example, the *npt*II gene, which confers resistance to kanamycin. The *Agrobacterium* which acts as host organism in this case should already contain a plasmid with the *vir* region. The latter is required for transferring the T-DNA to the plant cell. An *Agrobacterium* transformed in this way can be used for transforming plant cells. The use of T-DNA for transforming plant cells has been studied and described intensively (EP-A1 0 120 516; Hoekema 1985; An 1985; see also below).

Common binary vectors are based on "broad host range"-plasmids like pRK252 (Bevan 1984) or pTJS75 (Watson 1985) derived from the P-type plasmid RK2. Most of these vectors are derivatives of pBIN19 (Bevan 1984). Various binary vectors are known, some of which are commercially available such as, for example, pBI101.2 or pBIN19 (Clontech Laboratories, Inc. USA). Additional vectors were improved with regard to size and handling (*e.g.* pPZP; Hajdukiewicz 1994). Improved vector systems are described also in WO 02/00900.

In a preferred embodiment, *Agrobacterium* strains for use in the practice of the invention include octopine strains, *e.g.*, LBA4404 or agropine strains, *e.g.*, EHA101 or EHA105. Suitable strains of *A. tumefaciens* for DNA transfer are for example EHA101pEHA101 (Hood 1986), EHA105[pEHA105] (Li 1992), LBA4404[pAL4404] (Hoekema 1983), C58C1[pMP90] (Koncz 1986), and C58C1[pGV2260] (Deblaere 985). Other suitable strains are *Agrobacterium tumefaciens* C58, a nopaline strain. Other suitable strains are *A. tumefaciens* C58C1 (Van Laerebeke 1974), A136 (Watson et al. 1975) or LBA4011 (Klapwijk 1980). In a preferred embodiment, the *Agrobacterium* strain used to transform the plant tissue pre-cultured with the plant phenolic compound contains a L,L-succinamopine type Ti-plasmid, preferably disarmed, such as pEHA101. In another preferred embodiment, the *Agrobacterium* strain used to transform the plant tissue pre-cultured with the plant phenolic compound contains an octopine-type Ti-plasmid, preferably disarmed, such as pAL4404. Generally, when using octopine-type Ti-plasmids or helper plasmids, it is preferred that the *virF* gene be deleted or inactivated. In a preferred embodiment, the *Agrobacterium* strain used to transform the plant tissue pre-cultured with the plant phenolic compound such as acetosyringone. The method of the invention can also be used in combination with particular *Agrobacterium* strains, to further increase the transformation efficiency, such as *Agrobacterium* strains wherein the *vir* gene expression and/or induction thereof is altered due to the presence of mutant or chimeric *vir*A or *vir*G genes (*e.g.* Hansen 1994; Chen 1991; Scheeren-Groot 1994).

A binary vector or any other vector can be modified by common DNA recombination techniques, multiplied in *E. coli*, and introduced into *Agrobacterium* by *e.g.*, electroporation or other transformation techniques (Mozo 1991). *Agrobacterium* is grown and used as described in the art. The vector comprising *Agrobacterium* strain may, for example, be grown for 3 days on YP medium (5 g/L yeast extract, 10 g/L peptone, 5 g/L Nail, 15 g/L agar, pH 6.8) supplemented with the appropriate antibiotic (*e.g.,* 50 mg/L spectinomycin). Bacteria are collected with a loop from the solid medium and resuspended. For the purpose of this invention, *Agrobacterium* compatible vectors are pro-

vided by inserting site-specific recombination sites as described – for example – in the Examples.

After constructing a vector, the vector can be propagated in a host cell to synthesize nucleic acid molecules for the generation of a nucleic acid polymer. Vectors, often referred to as "shuttle vectors," are capable of replicating in at least two unrelated expression systems. To facilitate such replication, the vector should include at least two origins of replication, one effective in each replication system. Typically, shuttle vectors are capable of replicating in a eukaryotic system and a prokaryotic system. This enables detection of protein expression in eukaryotic hosts, the "expression cell type," and the amplification of the vector in the prokaryotic hosts, the "amplification cell type." As an illustration, one origin of replication can be derived from SV40, while another origin of replication can be derived from pBR322. Those of skill in the art know of numerous suitable origins of replication.

After constructing a vector, the vector is typically propagated in a host cell. Vector propagation is conveniently carried out in a prokaryotic host cell, such as *E. coli* or *Bacillus subtilus*. Suitable strains of *E. coli* include BL21(DE3), BL21(DE3)pLysS, BL21(DE3)pLysE, DB2, DB3.1, DH1, DH4I, DH5, DH5I, DH5IF, DH5IMCR, DH10B, DH10B/p3, DH11S, C600, HB101, JM101, JM105, JM109, JM110, K38, RR1, Y1088, Y1089, CSH18, ER1451, and ER1647 (see, for example, Brown (ed.), Molecular Biology Labfax (Academic Press 1991)). Suitable strains of *Bacillus subtilus* include BR151, YB886, MI119, MI120, and B170 (see, for example, Hardy 1985). Standard techniques for propagating vectors in prokaryotic hosts are well-known to those of skill in the art (see, for example, Ausubel 1995; Wu 1997).

### 3. THE SEQUENCES TO BE ASSESSED AS TRANSCRIPTION TERMINATORS

The sequences to be assessed using the method of the invention for their efficiency as transcription terminator sequences may be derived from various sources. In one embodiment of the invention sequences believed to function as transcription terminators may be assessed for their efficiency. Such sequences can be derived from for example regions downstream of the coding sequence of a gene (*e.g.*, comprising the region encoding the 3'-untranslated region and additional downstream genomic sequences), preferably from a region which is surrounding the end of the mRNA transcript. Various of such sequences can be derived from comparison of genomic and cDNA libraries. The corresponding nucleic acid sequences to be inserted into the insertion site of the screening vector or screening construct can be obtained for example by isolation from the corresponding host organism (by the various cloning methods known to the person skilled in the art, *e.g.*, by polymerase chain reaction employing appropriate primer oligonucleotides) or directly by DNA synthesis.

In a preferred embodiment the DNA molecules to be inserted for evaluation is a double-stranded, linear DNA molecule. The ends of said molecules may by blunt (*i.e.*, without 5'- and/or 3' overhangs) or "sticky" (*i.e.*, with 5'- and/ or 3' overhangs). Preferably, the ends of the DNA molecule may have overhangs which allow insertion into cleavage sites of restriction endonuclease to facilitate insertion into an insertion site. The length and molecular weight of the DNA molecule may vary. In a preferred embodiment the molecule has a size of about 50 to about 5,000 base pairs, preferably

from about 60 to about 2,000 base pairs, more preferably from about 70 to about 1,000 base pairs, most preferably from about 80 to about 500 base pairs.

Beside this educated approach (based on sequences for which some transcription termination efficiency can be presumed) in another preferred embodiment of the invention libraries of DNA sequences are screened to obtain efficient transcription terminator sequences. This embodiment does not require any previous sequence information and is based preferably solely on the phenotype of efficient transcription termination (which is difficult to correlate in practice with sequence information). The library of DNA sequences employed may be a synthetic library or a library of naturally occurring DNA molecules or a mixture of synthetic and naturally occurring DNA molecules. Preferably, the library of DNA molecules is a library of naturally occurring molecules, which may be derived from genomic DNA and/or cDNA of one or more organism. More preferably, the library is derived from the genomic DNA of an organism, preferably a plant organism.

In a preferred embodiment the DNA molecules of the DNA library are double-stranded, linear DNA molecules. The ends of said molecules may by blunt (*i.e.*, without 5'- and/or 3' overhangs) or "sticky" (*i.e.*, with 5'- and/ or 3' overhangs). The length and molecular weight of the DNA molecules of the library may vary. In a preferred embodiment the molecules have a size of about 50 to about 5,000 base pairs, preferably from about 60 to about 2,000 base pairs, more preferably from about 70 to about 1,000 base pairs, most preferably from about 80 to about 500 base pairs.

The library of DNA molecules may be derived from the genomic and/or cDNA by various means known to the person skilled in the art. For example, the library may be derived by random shearing of DNA of exhaustive or partial digestion with endonucleases. Preferably, the library is derived by exhaustive digestion with a restriction endonuclease, which has preferably a 4 base recognition site (like, *e.g.*, *Sau*3A). Following fragmentation (*e.g.*, by restriction), DNA molecules of the preferred molecular weight (as determined above) may be isolated by for example molecular weight exclusion chromatography (size exclusion chromatography using for example Superose™ columns, Amersham Bioscience, Inc.) or gel electrophoresis as known in the art (see for example Ellegren 1989).

In another preferred embodiment of the invention selected sequences can be assessed for their performance as transcription terminator sequences. Such sequences can be, for example, regions downstream of the coding sequence of a gene (*e.g.*, comprising the region encoding the 3'-untranslated region and additional downstream genomic sequences), preferably from a region which is surrounding the end of the mRNA transcript. Such sequences can be derived by *in silico* search of genome databases, such as for example of *Arabidopsis thaliana* or rice.

In one preferred embodiment, a partial or - preferably - entire plant genome (such as the rice or *Arabidopsis* genome) is screened for potential plant derived terminator candidates. The following criteria are used to identify and determine suitable candidates for transcription terminator sequences which may be further analyzed in the method of the invention:

1. Identification and/or isolation of intergenic regions between paired genes meeting predefined intergenic distance criteria. These genes may preferably have a head-to-tail orientation (*i.e.* transcription is running in the same direction), or - preferably - a tail-to-tail orientation (*i.e.* in opposite direction against each other). In the head-to-tail scenario the term "intergenic region" as used herein means the sequence in between (but excluding) the stop-codon of the "tail"-sequence and the start-codon of the "head" sequence, or – if known – the start of the promoter region of the "head" sequence. In the tail-to-tail orientation the term "intergenic region" as used herein means the sequence in between (but excluding) the two stop-codons of the coding sequence. Preferably, intergenic regions from paired genes in tail-to-tail orientation are identified which have a length from about 400 to 3,000 base pairs, preferably from about 700 to about 2,000 base pairs. Identification can be done by various means, including entire genome sequencing (*e.g.*, in case of previously unknown sequences) or *in silico* screening of already known sequences (such as the *Arabidopsis* or rice genome). Existing database can be employed for this purpose such as the most updated data from The Institute of Genome Research (TIGR; PUB_tigr_rice_genome_v4.nt (v03212003), PUB_tigr_rice_cds_Oct022003.nt, pubOSest0603 (ncbi)).

2. Identification and/or isolation of intergenic sequences which are flanked on both sides by genes having a high expression level. The term "high expression" or "high expression level" as used in this context means an expression level which is at least 5%, preferably at least 10%, more preferably at least 30%, most preferably at least 50% of the expression level of actin in the same mRNA source (*i.e.* cell or tissue). Expression may be judged by various means including but not limited to number of ESTs in a non-normalized EST library, Northern-blot analysis, RT-PCR *etc.* Low expression of one or both genes has been identified as an indicator for gene silencing by read-through transcription. Expression level can be profiled either by experiment (*e.g.*, *in vitro* or *in vivo* for example by using expression profiling by chip or microarray technology) or – preferably – *in silico* by simply counting the number of ESTs for each gene in non-normalized EST/cDNA-libraries which is indicative for expression level.

3. Identification and/or isolation of intergenic sequences which are flanked on both sides by genes having an expression pattern which is preferably independent from the expression pattern of the other paired gene. The term "independent expression pattern" in this context means – for example - that the expression of the first gene is different in its tissue and/or developmental regulation from the expression of the second gene. Dependency and correlation of expression patterns of paired genes has been identified as an indicator for read-through transcription. Expression profiles can be analyzed either by experiment by comparing expression level of said paired genes in various organs or tissues (*e.g.*, *in vitro* or *in vivo* for example by using expression profiling by chip technology) or – preferably – *in silico* by simply counting the number of ESTs for each gene in non-normalized EST/cDNA-libraries which is indicative for expression levels.

4. Identification and/or isolation of intergenic sequences which are flanked on one or – preferably - both sides by genes having a low variability in length of the mRNA transcript derived from said paired genes. Such variability is for example indicated by existence of more than one transcript end in EST, cDNA libraries or databases. Variability in transcript length has been identified as an indicator for low stringency in

transcription termination. Variability in transcript length can be analyzed either by experiment (*e.g.*, by RT PCR) or – preferably – *in silico* by simply analyzing the 3'-ends of EST or cDNA clones in the database.

While the intergenic localization (step 1) is a prerequisite, in a preferred embodiment of the invention each of the criteria 2, 3, 4, and the length of the intergenic region (part of criteria 1) for a certain intergenic sequence is resulting in a criteria score. Addition of said scores (which may be multiplied by certain weight-indicators reflecting the different impact of the criteria) is resulting in a final score which is indicative for the potential of the sequence as a transcription terminator and isolator (see below). This score and the potential can be verified by evaluating the sequence of said intergenic region in one or more screening systems of the invention. The highest weight is given to criteria 2 (high expression profile), followed by criteria 2 (independent expression profile), and criteria 3 (low variability in transcript length). The preferred length for the intergenic regions are indicated below, but seem to have more impact on handling (*i.e.* in later cloning and transformation procedures) than on functionality of said region.

An intergenic region identified thereby is not only suitable in mono-gene expression cassettes, but is especially suitable in multi-gene expression cassettes not only providing transcription termination for two genes in one sequence, but also allowing efficient "isolation" of said two expression cassettes by minimizing their interference by read-through transcription (thus providing an "isolator"), which has proven to be a serious problem especially in multi-gene expression constructs. The term "isolator" when referring to a sequence (which is preferably localized in between two expression cassettes) as used herein is intended to mean the capability of said sequence to minimize or prevent the influence of one expression cassette on transcription from the other expression cassette, thus isolating the two expression cassettes from each other. Preferred embodiments and additional information for carrying out this method for providing intergenic sequences is given in Example 1.2 below.

Accordingly, another embodiment of the invention is directed to a method for identification and/or isolation intergenic regions – preferably with high transcription termination and/or isolator potential – said method including at least one, preferably at least two, more preferably at least three, most preferably all of the following selection criteria

1. Identification and/or isolation or isolation of intergenic regions between paired genes meeting predefined intergenic distance criteria.
2. Identification and/or isolation of intergenic sequences which are flanked on both sides by genes having a high expression level.
3. Identification and/or isolation of intergenic sequences which are flanked on both sides by genes having an expression pattern which is preferably independent from the expression pattern of the other paired gene a high expression level.
4. Identification and/or isolation of intergenic sequences which are flanked on one or – preferably - both sides by genes having a low variability in length of the mRNA transcript derived from said paired genes.

Preferably the paired genes flanking the intergenic region have tail-to-tail orientation Preferably, intergenic regions from paired genes in tail-to-tail orientation are identified

which have a length from about 400 to 3,000 base pairs, preferably from about 700 to about 2,000 base pairs.

Thus, a preferred embodiment of the invention is related to a method for identification and/or isolation of intergenic regions with transcription termination potential said method including at least the steps of

a) identification and/or isolation or isolation of intergenic regions between paired genes having an intergenic distance of about 400 to 3,000 base pairs, and

b) identification and/or isolation of intergenic sequences which are flanked on both sides by genes having a high expression level.

More preferably said method further comprising the steps of

c) identification and/or isolation of intergenic sequences which are flanked on both sides by genes having an expression pattern which is preferably independent from the expression pattern of the other paired gene, and

d) identification and/or isolation of intergenic sequences which are flanked on one or – preferably - both sides by genes having a low variability in length of the mRNA transcript derived from said paired genes.

Most preferably, the intergenic region

a) is localized between genes which have a tail-to-tail localization (*i.e.* from which expression from said genes is directed in opposite direction against each other) and / or

b) has a length measured from the respective stop codons of the flanking genes from about 700 to about 2,000 base pairs.

Based on said method sequences from the rice genome were identified and found to be promising as transcription terminator sequences. Said sequences are described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, and 108.

Another embodiment of the invention is related to the use of said sequences to terminate transcription in a transgenic expression construct. More preferably is the use of said sequences as isolators in multi-gene expression constructs.

Another embodiment of the invention is related to a transgenic expression construct comprising in 5'-3'-direction

a) a promoter sequence functional in plants, and

b) a nucleic acid sequence of interest of to be expressed operably linked to said promoter a), and

c) at least one sequence selected from the group consisting of

i) the sequences described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, and 108, and

ii) the sequences having a homology of at least 60%, preferably 80%, more preferably 90%, most preferably 95% with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and

iii) the sequences hybridizing under low stringency conditions, preferably under high stringency conditions with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108 capable to terminate transcription in a plant cell or organism, and

iv) a fragment of at least 50 consecutive base pairs, preferably at least 100 consecutive base pairs, more preferably at least 250 consecutive base pairs, most preferably at least 500 consecutive base pairs of a sequence described under i), ii), and iii),

wherein said sequence c) is heterolog with respect to said promoter a) and/or said nucleic acid of interest b) and is mediating termination of expression of induced from said promoter a).

Another embodiment of the invention is related to a transgenic expression construct comprising at least two expression cassettes having a structure comprising in 5'-3'-direction

a1) a first promoter sequence functional in plants, and

b1) a first nucleic acid sequence of interest of to be expressed operably linked to said promoter a1), and

c) at least one sequence selected from the group consisting of

i) the sequences described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, and 108, and

ii) the sequences having a homology of at least 60%, preferably 80%, more preferably 90%, most preferably 95% with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and

iii) the sequences hybridizing under low stringency conditions, preferably under high stringency conditions with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and

iv) a fragment of at least 50 consecutive base pairs, preferably at least 100 consecutive base pairs, more preferably at least 250 consecutive base pairs, most preferably at least 500 consecutive base pairs of a sequence described under i), ii), and iii),

and,

b2) a second nucleic acid sequence of interest of to be expressed, and

a2) a second promoter sequence functional in plants operably linked to said nucleic acid sequence of interest b2),

wherein said sequence c) is heterolog with respect to at least one element selected from promoter a1), promoter a2), nucleic acid of interest b1) and nucleic acid of interest b2), and is mediating termination of expression of induced from said promoters a1) and a2).

Since no protein expression is caused from the above described transcription terminator sequences, a higher degree of variation is acceptable without changing the functionality.

The method of the invention can also be employed to identify regions responsible for transcription termination within larger sequences. This would allow to delete unnecessary sequences and to provide small sequences for transcription termination, which is an important goal in construction gene expression vectors. Large sequences leading to large vectors are linked to inefficient transformation and instability of constructs. Preferably, such identification can be realized by inserting fragments of a larger sequence into a screening vector or screening construct. Such fragments can be derived, for example, by nuclease mediated shorting of 5'- and/or 3'-ends of the larger sequence (by restrictions enzymes or unspecific nucleases such as *Bal*31). Corresponding methods are well known to the person skilled in the art.

The larger sequence for which one may seek to identify the essential region for transcription termination may for example be the natural region downstream of the coding sequence of the gene, which is the source for the promoter employed in the transgenic expression construct (*e.g.*, comprising the region encoding the 3'-untranslated region and additional downstream genomic sequences), preferably from a region which is surrounding the end of the mRNA transcript. It is advantageous to combine a promoter with its natural transcription terminator (and the heterogeneous sequence of interest in between) to obtain optimal expression results. While formerly either very long 3'-untranslated regions had to be employed to ensure efficient transcriptions termination, or laborious testing of shortened sequences had to be performed, the method of the present invention is allowing for fast and efficient restriction of a potential transcription terminator to its essential elements.

## 4. INSERTION OF THE DNA MOLECULES INTO THE SCREENING VECTOR

The DNA molecules to be assessed for their transcription termination efficiency may be inserted into the Screening Vector by various means. Preferably, the insertion is realized by one or more methods selected from the group consisting of:

a) Insertion into a restriction site: Various sequence specific endonuclease are known to the person skilled in art which can be employed for carrying out the method of the invention. Suitable endonuclease may be for example type II restriction endonucleases or artificial (*e.g.*, chimeric) nucleases. Preferred are restriction endonucleases which are chosen in a way that only the insertion site is cleaved by said restriction endonuclease. Such restriction endonuclease may preferably include rare cutting endonucleases which have a recognition site of at least 8 base pairs (such as for example *Not*I) or even homing endonucleases, which have very long recognition sequences (Belfort 1997; Jasin 1996; Internet: http://rebase.neb.com/rebase/ re-

base.homing.html; Roberts 2001). Examples for preferred homing endonucleases include but are not limited to F-*Sce*I,I-*Ceu*I,I-*Chu*I, I-*Dmo*I, I-*Cpa*I, I-*Cpa*II, I-*Cre*I, I-*Csm*I, F-*Tev*I, F-*Tev*II, I-*Tev*I, I-*Tev*II, I-*Ani*I, I-*Cvu*I, I-*Lla*I, I-*Nan*I, I-*Mso*I, I-*Nit*I, I-*Nja*I, I-*Pak*I, I-*Por*I, I-*Ppo*I, I-*Sca*I, I-*Ssp*6803I, PI-*Pko*I, PI-*Pko*II, PI-*Psp*I, PI-*Tfu*I, PI-*Tli*I. Most preferred are I-*Ceu*I, I-*Sce*I, I-*Ppo*I, PI-*Psp*I, and PI-*Sce*I.

b) Insertion into a recombination site: In a preferred embodiment of the invention, the insertion of DNA segments into the insertion site of the screening construct or screening vector is achieved by the use of recombination proteins, including recombinases and associated co-factors and proteins. Numerous recombination systems from various organisms can also be used, based on the teaching and guidance provided herein. See, *e.g.*, Hoess 1986; Abremski 1986; Campbell, 1992; Qian 1992; Araki 1992). Many of these belong to the integrase family of recombinases (Argos 1986). Perhaps the best studied of these are the Integrase/att system from bacteriophage λ (Landy 1993), the Cre/loxP system from bacteriophage P1 (Hoess 1990), and the FLP/FRT system from the *Saccharomyces cerevisiae* 2μ circle plasmid (Broach 1982)). Detailed method for recombinase mediated cloning, appropriate recombination sites (to be employed as insertion sites), and corresponding recombinases are described *e.g.*, in US 5,888,732, hereby incorporated entirely by reference. A preferred system is the Gateway™ cloning system (Invitrogen, Inc.). Corresponding ready-to-use mixture of lambda integrase with its corresponding co-factors can be obtained from Invitrogen Inc. (Gateway™ LR Clonase™ Plus enzyme).

Also procedures comprising combination of both method can be employed. However, it is not essential that the sequence to be assessed is inserted directly (*i.e.* in an one-step cloning procedure) into a quasi ready-to-go screening construct or screening vector. It may for example – first be linked to the additional sequence and then inserted into an appropriate construct or vector thus constituting the final screening construct or screening vector to be employed in the evaluation procedure. In principle, the ways and possibilities to assembly the various parts of a screening construct or screening vector are uncountable but well known and established to the person skilled in the art.

For the purpose of insertion into the insertion site the DNA sequence to be inserted may be linked to adapters providing the appropriate recognition sequences for restriction endonuclease or recombinase, respectively. However, in the case of restriction endonucleases adapters are not required in cases where a digestion of genomes is employed as a library of DNA sequences. Here the restriction enzyme employed for the digest should create DNA ends compatible with those at the cleaved insertion site.

## 5. THE *IN VITRO* SCREENING SYSTEM

When performed as an *in vitro* screening system, the expression of the additional sequences (which may preferably be located downstream (*i.e.* in 3'-direction) of the insertion site) may be – for example - easily detected at the RNA levels using sensitive fluorescence probes that recognize single strand nucleotides. Such features are also to be understood as readily detectable characteristics. In case an efficient transcription terminator sequence is inserted in front of these sequences a reduced, preferably no significant or at all observable signal will be obtained.

Within the *in vitro* screening system, transcription of sequences located downstream of the transcription termination sequence inserted into the insertion site can be detected at the RNA levels using commercially available *in vitro* transcription systems (such as wheat germ nuclear extracts, HeLa nuclear extracts, rabbit reticulocyte extracts, or nuclear extracts from plant of interest) preferably in combination with single strand recognizing florescence probes (*e.g.* beacon probes). Various suitable *in vitro* transcription /translation systems are known in the art and commercially available (*e.g.,* ActivePro™, PROTEINscript™ II, Retic Lysate IVT™ (treated) and Retic Lysate IVT™-96, Wheat Germ IVT™; all available from Ambion, Inc., Austin, USA). In this case no plant transformation is involved. In consequence, the screening construct or screening vector can be constructed on a simple base (*e.g.*, pUC based). Preferably, individual screening constructs or screening vectors comprising different transcription termination sequences are placed in 96 well plates for *in vitro* transcription. The fluorescent probe hybridizes when read through occurs. The tighter transcription termination occurs, the less fluorescent products in the read through region are detected. The amounts of read through products can be normalized by the expression of sequences located upstream of the transcription termination sequences but still under control of the promoter.

## 6. THE *IN VIVO* SCREENING SYSTEM

In a preferred embodiment, the method of the invention is realized *in vivo*, preferably in the target organism in which an efficient transcription terminator is sought for. The *in vivo* screening system allows for evaluation of multiple DNA sequences for their performance as transcription terminator sequences in parallel. Thus, a library of DNA sequences can be employed and inserted into the screening construct or screening vector yielding a library of screening constructs or screening vectors comprising various different DNA sequences. Said library of screening constructs or screening vectors is inserted into cells or organisms in a way that each individual cell or organism preferably comprises only one screening constructs or screening vectors of said library (comprising one specific DNA sequence to be assessed for the transcription termination capability). In consequence – as described below in more detail - this preferred embodiment does not necessarily require the sorting of the various screening constructs or screening vectors prior to evaluation for the transcription termination capability, which makes the method even more efficient. Thus, in a preferred embodiment the method of the invention therefore relates to a method for identification and isolation of transcription termination sequences for comprising the steps of:

i) providing a screening construct or screening vector comprising
   a) a promoter sequence, and
   b) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences, and
   c) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic,

   wherein insertion of an efficient transcription terminator into said insertion site changes expression of said additional sequences by said promoter sequence in comparison to no insertion, and

ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and

iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector, and

iv) introducing said screening construct or screening vector with said inserted DNA sequences into cells or organisms suitable to induce expression from said promoter sequence, and

v) identifying and/or selecting cells or organisms with a changed readily detectable characteristic in comparison to no insertion, and

vi) isolating the inserted DNA sequences from said identified and/or selected screening construct or screening vector for use as transcription termination sequences and – optionally – determining their sequence.

All the above specified preferred variations (such as Variation A, B, or C) can also be advantageously combined with said *in vivo* system. In the *in vivo* screening systems of the invention, the expression of the sequence located downstream (at the 3'-end) of the insertion site will preferably cause an easily detectable phenotype. "Causing" includes both initiating or suppressing an easily detectable phenotype. For example, the sequence located downstream (at the 3'-end) of the insertion site may either code for a phenotype causing protein, or it may code for RNA (*e.g.*, antisense or double stranded RNA) which causes suppression of expression of a phenotype causing protein. Multiple examples are given above.

In an preferred embodiment of the *in vivo* screening systems of the invention, the expression of the sequence located downstream (at the 3'-end) of the insertion site will cause a phenotype which is inhibiting growth, propagation and/or or regeneration of said cells or organisms (*e.g.*, plant cells or plants), and which is therefore understood within the context of this invention to be "toxic" to said cells and/organisms (*e.g.*, plant cells or plants). In consequence, only cells (or organisms) will survive if a tight transcription termination sequence is inserted in front of said toxic phenotype causing sequence thereby preventing expression of this growth, propagation and/or or regeneration inhibiting sequences. The surviving cells can be isolated and the transcription terminator sequence can be identified and isolated, *e.g.*, by amplification using PCR followed by sequencing.

For conducting the screening in the *in vivo* system the screening construct or screening vector in transformed preferably into a cell, tissue or organism. The generation of a transformed organism or a transformed cell requires introducing the DNA in question into the host cell in question. A multiplicity of methods is available for this procedure, which is termed transformation (see also Keown 1990). For example, the DNA can be introduced directly by microinjection or by bombardment via DNA-coated microparticles. Also, the cell can be permeabilized chemically, for example using polyethylene glycol, so that the DNA can enter the cell by diffusion. The DNA can also be introduced by protoplast fusion with other DNA-containing units such as minicells, cells, lysosomes or liposomes. Another suitable method of introducing DNA is electroporation, where the cells are permeabilized reversibly by an electrical pulse.

The host cell or organism can be any prokaryotic or eukaryotic organism. Preferred are mammalian cells, non-human mammalian organism, plant cells and plant organisms as defined above.

The screening construct or screening vector of the invention is preferably introduced into a eukaryotic cell. It may be preferably inserted into the genome (*e.g.*, plastids or chromosomal DNA) but may also be exist extra-chromosomal or epichromosomal. Preferred eukaryotic cells are mammalian cell, fungal cell, plant cell, insect cell, avian cell, and the like. Examples of suitable mammalian host cells include African green monkey kidney cells (Vero; ATCC CRL 1587), human embryonic kidney cells (293-HEK; ATCC CRL 1573), baby hamster kidney cells (BHK-21, BHK-570, ATCC CRL 8544, ATCC CRL 10314), canine kidney cells (MDCK; ATCC CCL 34), Chinese hamster ovary cells (CHO-K1; ATCC CCL61; CHO DG44 (Chasin 1986), rat pituitary cells (GH1; ATCC CCL82), HeLa S3 cells (ATCC CCL2.2), rat hepatoma cells (H-4-II-E; ATCC CRL 1548) SV40-transformed monkey kidney cells (COS-1; ATCC CRL 1650) and murine embryonic cells (NIH-3T3; ATCC CRL 1658).

A screening construct or screening vector can be introduced into host cells using a variety of standard techniques including calcium phosphate transfection, liposome-mediated transfection, microprojectile-mediated delivery, electroporation, and the like. Transfected cells can be selected and propagated to provide recombinant host cells that comprise the gene of interest stably integrated in the host cell genome.

The screening vector may be a baculovirus expression vector to be employed in a baculovirus system. The baculovirus system provides an efficient means to introduce cloned genes of interest into insect cells. Suitable expression vectors are based upon the *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV), and contain well-known promoters such as Drosophila heat shock protein (hsp) 70 promoter, *Autographa californica* nuclear polyhedrosis virus immediate-early gene promoter (ie-1) and the delayed early 39K promoter, baculovirus p10 promoter, and the Drosophila metallothionein promoter. A second method of making recombinant baculovirus utilizes a transposon-based system (Luckow 1993). This system, which utilizes transfer vectors, is sold in the BAC-to-BAC kit (Life Technologies, Rockville, Md.). This system utilizes a transfer vector, PFASTBAC (Life Technologies) containing a Tn7 transposon to move the DNA encoding the polypeptide of interest into a baculovirus genome maintained in E. coli as a large plasmid called a "bacemid" (see, Hill-Perkins 1990; Bonning 1994; and Chazenbalk 1995). In addition, transfer vectors can include an in-frame fusion with DNA encoding an epitope tag at the C- or N-terminus of the expressed polypeptide, for example, a Glu-Glu epitope tag (Grussenmeyer 1985). Using a technique known in the art, a transfer vector containing a gene of interest is transformed into *E. coli*, and screened for bacmids, which contain an interrupted lacZ gene indicative of recombinant baculovirus. The bacmid DNA containing the recombinant baculovirus genome is then isolated using common techniques. The recombinant virus or bacinid is used to transfect host cells. Suitable insect host cells include cell lines derived from IPLB-Sf-21, a Spodoptera frugiperda pupal ovarian cell line, such as Sf9 (ATCC CRL 1711), Sf21AE, and Sf21 (Invitrogen Corporation; San Diego, Calif.), as well as Drosophila Schneider-2 cells, and the HIGH FIVE ™ cell line (Invitrogen) derived from *Trichoplusia ni* (US 5,300,435). Commercially available serum-free media can be used to grow and to maintain the cells. Suitable media are Sf900 II™ (Life Technologies) or ESF 921™(Expression Systems) for the Sf9 cells; and Ex-cellO405™ (JRH Biosciences, Lenexa, Kans.) or Express Five™ (Life Technologies) for the *T. ni* cells. When recombinant virus is used, the cells are typically grown up from an inoculation density of approximately 2-5.times.10.sup.5 cells to a density of 1-2.times.10.sup.6 cells at which

time a recombinant viral stock is added at a multiplicity of infection of 0.1 to 10, more typically near 3. Established techniques for the baculovirus systems are provided by Bailey 1991, Patel 1995, Ausubel 1995 (at pages 16-37 to 16-57), Richardson 1995, and by Lucknow, 1996.

Fungal cells, including yeast cells, can also be used as host cells for transformation with the screening construct or screening vector of the invention. Yeast species of particular interest in this regard include *Saccharomyces cerevisiae, Pichia pastoris, and Pichia methanolica*. Suitable promoters for expression in yeast include promoters from GAL1 (galactose), PGK (phosphoglycerate kinase), ADH (alcohol dehydrogenase), AOX1 (alcohol oxidase), HIS4 (histidinol dehydrogenase), and the like. Many yeast cloning vectors have been designed and are readily available to be employed. These vectors include YIp-based vectors, such as YIp5, YRp vectors, such as YRp17, YEp vectors such as YEp13 and YCp vectors, such as YCp19. Methods for transforming S. cerevisiae cells with exogenous DNA and producing recombinant polypeptides therefrom are disclosed by, for example, US 4,599,311, US 4,931,373, US 4,870,008, US 5,037,743, and US 4,845,075. Transformed cells are selected by phenotype determined by the selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient (*e.g.*, leucine). An illustrative vector system for use in Saccharomyces cerevisiae is the POTI vector system (US 4,931,373), which allows transformed cells to be selected by growth in glucose-containing media. Additional suitable promoters and terminators for use in yeast include those from glycolytic enzyme genes (see, *e.g.*, US 4,599,311, US 4,615,974, and US 4,977,092) and alcohol dehydrogenase genes. See also US 4,990,446, 5,063,154, 5,139,936, and 4,661,454.

Transformation systems for other yeasts, including *Hansenula polymorpha, Schizosaccharomyces pombe, Kluyveromyces lactis, Kluyveromyces fragilis, Ustilago maydis, Pichia pastoris, Pichia methanolica, Pichia guillermondii* and *Candida maltosa* are known in the art. See, for example, Gleeson 1986, and US 4,882,279. *Aspergillus* cells may be utilized according to the methods of McKnight *et al.* (US 4,935,349). Methods for transforming *Acremonium chrysogenum* are disclosed (US 5,162,228). Methods for transforming *Neurospora* are disclosed (US 4,486,533).

For example, the use of *Pichia methanolica* as host for the production of recombinant proteins is disclosed (US 5,716,808, US 5,736,383, Raymond 1998, WO 97/17450, WO 97/17451, WO 98/02536, and WO 98/02565). DNA molecules for use in transforming *P. methanolica* will commonly be prepared as double-stranded, circular plasmids, which are preferably linearized prior to transformation. For polypeptide production in *P. methanolica*, the promoter and terminator in the plasmid can be that of a *P. methanolica* gene, such as a *P. methanolica* alcohol utilization gene (AUG1 or AUG2). Other useful promoters include those of the dihydroxyacetone synthase (DHAS), formnate dehydrogenase (FMD), and catalase (CAT) genes. To facilitate integration of the DNA into the host chromosome, it is preferred to have the entire expression segment of the plasmid flanked at both ends by host DNA sequences. For large-scale, industrial processes where it is desirable to minimize the use of methanol host cells can be used in which both methanol utilization genes (AUG1 and AUG2) are deleted. For production of secreted proteins, host cells can be used that are deficient in vacuolar protease genes (PEP4 and PRB1). Electroporation is used to facilitate the introduction of a plasmid containing DNA encoding a polypeptide of interest into *P. methanolica cells. P. metha-*

*nolica cells* can be transformed by electroporation using an exponentially decaying, pulsed electric field having a field strength of from 2.5 to 4.5 kV/cm, preferably about 3.75 kV/cm, and a time constant (t) of from 1 to 40 milliseconds, most preferably about 20 milliseconds.

Standard methods for introducing nucleic acid molecules into bacterial, yeast, insect, mammalian, and plant cells are provided, for example, by Ausubel (1995). General methods for expressing and recovering foreign protein produced by a mammalian cell system are provided by, for example, in Etcheverry 1996. Established methods for isolating recombinant proteins from a baculovirus system are described (Richardson 1995).

Especially preferred in transfer of the screening construct or screening vector into plant cells, tissues and/or organism. Methods for introduction of a transgenic expression construct or vector into plant tissue may include but are not limited to, *e.g.*, electroinjection (Nan 1995; Griesbach 1992); fusion with liposomes, lysosomes, cells, minicells or other fusible lipid-surfaced bodies (Fraley 1982); polyethylene glycol (Krens 1982); chemicals that increase free DNA uptake; transformation using virus, and the like. Furthermore, the biolistic method with the gene gun, electroporation, incubation of dry embryos in DNA-containing solution, and microinjection may be employed.

Protoplast based methods can be employed (*e.g.*, for rice), where DNA is delivered to the protoplasts through liposomes, PEG, or electroporation (Shimamoto 1989; Datta 1990b). Transformation by electroporation involves the application of short, high-voltage electric fields to create ."pores" in the cell membrane through which DNA is taken-up. These methods are – for example - used to produce stably transformed monocotyledonous plants (Paszkowski 1984; Shillito 1985; Fromm 1986) especially from rice (Shimamoto 1989; Datta 1990b; Hayakawa 1992).

Particle bombardment or "biolistics" is a widely used method for the transformation of plants, especially monocotyledonous plants. In the "biolistics" (microprojectile-mediated DNA delivery) method microprojectile particles are coated with DNA and accelerated by a mechanical device to a speed high enough to penetrate the plant cell wall and nucleus (WO 91/02071). The foreign DNA gets incorporated into the host DNA and results in a transformed cell. There are many variations on the "biolistics" method (Sanford 1990; Fromm 1990; Christou 1988; Sautter 1991). The method has been used to produce stably transformed monocotyledonous plants including rice, maize, wheat, barley, and oats (Christou 1991; Gordon-Kamm 1990; Vasil 1992; Wan 1994).

In addition to these "direct" transformation techniques, transformation can also be affected by bacterial infection by means of *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*. These strains contain a plasmid (Ti or Ri plasmid) which is transferred to the plant following *Agrobacterium* infection. Part of this plasmid, termed T-DNA (transferred DNA), is integrated into the genome of the plant cell (see above for description of vectors). To transfer the DNA to the plant cell, plant explants are cocultured with a transgenic *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*. Starting from infected plant material (for example leaf, root or stem sections, but also protoplasts or suspensions of plant cells), intact plants can be generated using a suitable medium which may contain, for example, antibiotics or biocides for selecting transformed cells.

The plants obtained can then be screened for the presence of the DNA introduced, in this case the expression construct according to the invention. As soon as the DNA has integrated into the host genome, the genotype in question is, as a rule, stable and the insertion in question is also found in the subsequent generations. As a rule, the expression construct integrated contains a selection marker which imparts a resistance to a biocide (for example a herbicide) or an antibiotic such as kanamycin, G 418, bleomycin, hygromycin or phosphinotricin and the like to the transformed plant. The selection marker permits the selection of transformed cells from untransformed cells (McCormick 1986). The plants obtained can be cultured and hybridized in the customary fashion. Two or more generations should be grown in order to ensure that the genomic integration is stable and hereditary. The abovementioned methods are described in detail in the relevant art (for example, in Jenes 1993, and in Potrykus 1991).

One of skill in the art knows that the efficiency of transformation by *Agrobacterium* may be enhanced by using a number of methods known in the art. For example, the inclusion of a natural wound response molecule such as acetosyringone (AS) to the *Agrobacterium* culture has been shown to enhance transformation efficiency with *Agrobacterium tumefaciens* (Shahla 1987). Alternatively, transformation efficiency may be enhanced by wounding the target tissue to be transformed. Wounding of plant tissue may be achieved, for example, by punching, maceration, bombardment with microprojectiles, *etc.* (see, *e.g.*, Bidney 1992).

A number of other methods have been reported for the transformation of plants (especially monocotyledonous plants) including, for example, the "pollen tube method" (WO 93/18168; Luo 1988), macro-injection of DNA into floral tillers (Du 1989; de la Pena 1987), injection of *Agrobacterium* into developing caryopses (WO 00/63398), and tissue incubation of seeds in DNA solutions (Topfer 1989). Direct injection of exogenous DNA into the fertilized plant ovule at the onset of embryogenesis was disclosed in WO 94/00583. WO 97/48814 disclosed a process for producing stably transformed fertile wheat and a system of transforming wheat via *Agrobacterium* based on freshly isolated or pre-cultured immature embryos, embryogenic callus and suspension cells.

It may be desirable to target the nucleic acid sequence of interest to a particular locus on the plant genome. Site-directed integration of the nucleic acid sequence of interest into the plant cell genome may be achieved by, for example, homologous recombination using *Agrobacterium*-derived sequences. Generally, plant cells are incubated with a strain of *Agrobacterium* which contains a targeting vector in which sequences that are homologous to a DNA sequence inside the target locus are flanked by *Agrobacterium* transfer-DNA (T-DNA) sequences, as previously described (US 5,501,967, the entire contents of which are herein incorporated by reference). One of skill in the art knows that homologous recombination may be achieved using targeting vectors which contain sequences that are homologous to any part of the targeted plant gene, whether belonging to the regulatory elements of the gene, or the coding regions of the gene. Homologous recombination may be achieved at any region of a plant gene so long as the nucleic acid sequence of regions flanking the site to be targeted is known.

Where homologous recombination is desired, the targeting vector used may be of the replacement- or insertion-type (US 5,501,967). Replacement-type vectors generally contain two regions which are homologous with the targeted genomic sequence and

which flank a heterologous nucleic acid sequence, *e.g.*, a selectable marker gene sequence. Replacement-type vectors result in the insertion of the selectable marker gene which thereby disrupts the targeted gene. Insertion-type vectors contain a single region of homology with the targeted gene and result in the insertion of the entire targeting vector into the targeted gene.

Transformed cells, *i.e.* those which contain the introduced DNA integrated into the DNA of the host cell, can be selected from untransformed cells if a selectable marker is part of the introduced DNA. A selection marker gene may confer positive or negative selection.

A positive selection marker gene may be used in constructs for random integration and site-directed integration. Positive selection marker genes include antibiotic resistance genes, and herbicide resistance genes and the like. Transformed cells which express such a marker gene are capable of surviving in the presence of concentrations of the antibiotic or herbicide in question which kill an untransformed wild type. Examples are the bar gene, which imparts resistance to the herbicide phosphinotricin (bialaphos; Vasil 1992; Weeks 1993; Rathore 1993), the *npt*II gene, which imparts resistance to kanamycin, the *hpt* gene, which imparts resistance to hygromycin, or the EPSP gene, which imparts resistance to the herbicide glyphosate, geneticin (G-418) (aminoglycoside) (Nehra 1994), glyphosate (Della-Cioppa *et al.* 1987) and the ALS gene (chlorsulphuron resistance). Further preferred selectable and screenable marker genes are disclosed above.

A negative selection marker gene may also be included in the constructs. The use of one or more negative selection marker genes in combination with a positive selection marker gene is preferred in constructs used for homologous recombination. Negative selection marker genes are generally placed outside the regions involved in the homologous recombination event. The negative selection marker gene serves to provide a disadvantage (preferably lethality) to cells that have integrated these genes into their genome in an expressible manner. Cells in which the targeting vectors for homologous recombination are randomly integrated in the genome will be harmed or killed due to the presence of the negative selection marker gene. Where a positive selection marker gene is included in the construct, only those cells having the positive selection marker gene integrated in their genome will survive. The choice of the negative selection marker gene is not critical to the invention as long as it encodes a functional polypeptide in the transformed plant cell. The negative selection gene may for instance be chosen from the aux-2 gene from the Ti-plasmid of *Agrobacterium*, the tk-gene from SV40, cytochrome P450 from *Streptomyces griseolus*, the *Adh* gene from Maize or *Arabidopsis*, *etc.* Any gene encoding an enzyme capable of converting a substance which is otherwise harmless to plant cells into a substance which is harmful to plant cells may be used. Further preferred negative selection markers are disclosed above.

However, insertion of an expression cassette or a vector into the chromosomal DNA can also be demonstrated and analyzed by various other methods (not based on selection marker) known in the art like including, but not limited to, restriction mapping of the genomic DNA, PCR-analysis, DNA-DNA hybridization, DNA-RNA hybridization, DNA sequence analysis and the like. More specifically such methods may include *e.g.*, PCR analysis, Southern blot analysis, fluorescence *in situ* hybridization (FISH), and *in situ*

PCR.

As soon as a transformed plant cell has been generated, an intact plant can be obtained using methods known to the skilled worker. Accordingly, the present invention provides transgenic plants. The transgenic plants of the invention are not limited to plants in which each and every cell expresses the nucleic acid sequence of interest under the control of the promoter sequences provided herein. Included within the scope of this invention is any plant which contains at least one cell which expresses the nucleic acid sequence of interest (*e.g.*, chimeric plants). It is preferred, though not necessary, that the transgenic plant comprises the nucleic acid sequence of interest in more than one cell, and more preferably in one or more tissue.

Once transgenic plant tissue which contains an expression vector has been obtained, transgenic plants may be regenerated from this transgenic plant tissue using methods known in the art. The term "regeneration" as used herein, means growing a whole plant from a plant cell, a group of plant cells, a plant part or a plant piece (*e.g.*, from a protoplast, callus, protocorm-like body, or tissue part).

Species from the following examples of genera of plants may be regenerated from transformed protoplasts: *Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciohorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Pisum, Lolium, Zea, Triticum, Sorghum*, and *Datura.*

For regeneration of transgenic plants from transgenic protoplasts, a suspension of transformed protoplasts or a Petri plate containing transformed explants is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, somatic embryo formation can be induced in the callus tissue. These somatic embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and plant hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. These three variables may be empirically controlled to result in reproducible regeneration.

Plants may also be regenerated from cultured cells or tissues. Dicotyledonous plants which have been shown capable of regeneration from transformed individual cells to obtain transgenic whole plants include, for example, apple (*Malus pumila*), blackberry (*Rubus*), Blackberry/raspberry hybrid (*Rubus*), red raspberry (*Rubus*), carrot (*Daucus carota*), cauliflower (*Brassica oleracea*), celery (*Apium graveolens*), cucumber (*Cucumis sativus*), eggplant (*Solanum melongena*), lettuce (*Lactuca sativa*), potato (*Solanum tuberosum*), rape (*Brassica napus*), wild soybean (*Glycine canescens*), strawberry (*Fragaria ananassa*), tomato (*Lycopersicon esculentum*), walnut (*Juglans regia*), melon (*Cucumis melo*), grape (*Vitis vinifera*), and mango (*Mangifera indica*). Monocotyledonous plants which have been shown capable of regeneration from transformed individual cells to obtain transgenic whole plants include, for example, rice (*Oryza sativa*), rye

(*Secale cereale*), and maize (*Zea mays*).

In addition, regeneration of whole plants from cells (not necessarily transformed) has also been observed in: apricot (*Prunus armeniaca*), asparagus (*Asparagus officinalis*), banana (hybrid *Musa*), bean (*Phaseolus vulgaris*), cherry (hybrid *Prunus*), grape (*Vitis vinifera*), mango (*Mangifera indica*), melon (*Cucumis melo*), ochra (*Abelmoschus esculentus*), onion (hybrid *Allium*), orange (*Citrus sinensis*), papaya (*Carrica papaya*), peach (*Prunus persica*), plum (*Prunus domestica*), pear (*Pyrus communis*), pineapple (*Ananas comosus*), watermelon (*Citrullus vulgaris*), and wheat (*Triticum aestivum*).

The regenerated plants are transferred to standard soil conditions and cultivated in a conventional manner. After the expression vector is stably incorporated into regenerated transgenic plants, it can be transferred to other plants by vegetative propagation or by sexual crossing. For example, in vegetatively propagated crops, the mature transgenic plants are propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. In seed propagated crops, the mature transgenic plants are self-crossed to produce a homozygous inbred plant which is capable of passing the transgene to its progeny by Mendelian inheritance. The inbred plant produces seed containing the nucleic acid sequence of interest. These seeds can be grown to produce plants that would produce the selected phenotype. The inbred plants can also be used to develop new hybrids by crossing the inbred plant with another inbred plant to produce a hybrid.

Confirmation of the transgenic nature of the cells, tissues, and plants may be performed by PCR analysis, antibiotic or herbicide resistance, enzymatic analysis and/or Southern blots to verify transformation. Progeny of the regenerated plants may be obtained and analyzed to verify whether the transgenes are heritable. Heritability of the transgene is further confirmation of the stable transformation of the transgene in the plant. The resulting plants can be bred in the customary fashion. Two or more generations should be grown in order to ensure that the genomic integration is stable and hereditary. Corresponding methods are described, (Jenes 1993; Potrykus 1991).

**7. Conduction the Screening, Isolation and Use of the Transcription Terminator Sequences**

Once one or more DNA sequences or even a library of sequences to be assessed for their transcription termination efficiency was inserted into the screening construct or screening vector these vectors are submitted to the appropriate *in vitro* or *in vivo* screening system.

The readily detectable characteristic or the change thereof can be monitored by various means well known to the person skilled in the art depending on the additional sequence employed. The "output" of the screening system (*i.e.* the number of different transcription terminator sequences) and their efficiency can be controlled by setting certain cut-off limits. For example a certain intensity of color or fluorescence (in case the characteristic is a color), a certain resistance against a toxic compound (in case the characteristic is a resistance).

Screening constructs, screening vectors, or cells or organisms comprising those, derived from the screening process can be employed to isolate and analyze the transcrip-

tion termination sequences comprised therein. Isolation can be done by various means including but not limited to PCR mediated amplification of the sequence inserted into the insertion site using primers specific for the known regions flanking said insertion site.

The isolated transcription terminator sequence can be used for various purposes in biotechnology, preferably in constructing gene expression constructs which require a tight transcription termination control *i.e.* a low read-through frequency. Such expression cassettes (consisting for example in 5'/3'-direction of a promoter, a gene of interest, and the isolated transcription termination sequence) can be produced by means of customary recombination and cloning techniques as are described (for example, in Maniatis 1989; Silhavy 1984; and in Ausubel 1987). The person skilled in the art is aware of numerous sequences which may be utilized as gene of interest in this context, *e.g.* to increase quality of food and feed, to produce chemicals, fine chemicals or pharmaceuticals (*e.g.*, vitamins, oils, carbohydrates; Dunwell 2000), conferring resistance to herbicides, or conferring male sterility. Furthermore, growth, yield, and resistance against abiotic and biotic stress factors (like *e.g.*, fungi, viruses, nematodes, or insects) may be enhanced. Advantageous properties may be conferred either by over-expressing proteins or by decreasing expression of endogenous proteins by *e.g.*, expressing a corresponding antisense (Sheehy 1988; US 4,801,340; Mol 1990) or double-stranded RNA (Matzke 2000; Fire 1998; Waterhouse 1998; WO 99/32619; WO 99/53050; WO 00/68374; WO 00/44914; WO 00/44895; WO 00/49035; WO 00/63364).

It will be understood by one of ordinary skill in the relevant arts that other suitable modifications and adaptations to the methods and applications described herein are readily apparent and may be made without departing from the scope of the invention or any embodiment thereof. Having now described the present invention in detail, the same will be more clearly understood by reference to the following examples, which are included herewith for purposes of illustration only and are not intended to be limiting of the invention.

**Sequences**

1. SEQ ID NO: 1 Binary expression vector Lo546b-pSUN1-R4-Lo484::Lo376::Lo522b

2. SEQ ID NO: 2 Binary expression vector Lo546a-pSUN1-R4-Lo484::Lo376::Lo522a

3. SEQ ID NO: 3 Nucleic acid construct Lo522b-pENTR-C1-STPT-*nptII*-IRnos

4. SEQ ID NO: 4 Nucleic acid construct Lo522a-pENTR-C1-STPT-*nptII*-IRnos

5. SEQ ID NO: 5 Binary expression vector Lo523b-pSUN1-R4-Lo484::Lo376::Lo503b

6. SEQ ID NO: 6 Binary expression vector Lo523a-pSUN1-R4-Lo484::Lo376::Lo503a

7. SEQ ID NO: 7 Nucleic acid construct Lo503b-pENTR-C1-STPT-*nptII*-IRnos

8. SEQ ID NO: 8 Nucleic acid construct Lo503a-pENTR-C1-STPT-*nptII*-IRnos

9. SEQ ID NO: 9 Nucleic acid construct Lo484-pENTR-A1-inv-35s-GFP-E9

10. SEQ ID NO: 10-46: Nucleic acid sequence from rice (*Oryza sativa sbsp. japonica*) encoding sequences suitable as transcription terminators and expression cassette isolators.

11. SEQ ID NO: 47: Primer 1(*Sac*I, *Avr*II, *Spe*I, OCS 5')
5'-CG GAGCTC CCTAGG ACTAGT tcgaccggcatgccc-3'

12. SEQ ID NO: 48 Primer 2 (*Not*I, OCS 3')
5'- CC GCGGCCGC agcttggacaatcag-3'

13. SEQ ID NO: 49 Primer 3 (*Avr*II, *Xma*I, *Rsr*II, $Lu^F$ 5')
5'-CG CCTAGG CCCGGG CGGACCG cattaagaagggccc-3'

14. SEQ ID NO: 50 Primer 4 (*Spe*I $Lu^F$ 3')
5'-CG ACTAGT agagagttctcagagc-3'

15. SEQ ID NO: 51 Primer 5 (*Rsr*II, *Bsp*EI, target gene seq 5')
5' CG CGGACCG TCCGGA-N-3'
[N represents a gene-specific sequence of preferably 10 to 20 bases]

16. SEQ ID NO: 52 Primer 6 (*Spe*I, *Age*I, target gene seq 3')
5' CG ACTAGT ACCGGT-N-3'
[N represents a gene-specific sequence of preferably 10 to 20 bases]

17. SEQ ID NO: 53 pTOI3

18. SEQ ID NO: 54 pTOI4

19. SEQ ID NO: 55 Oligonucleotideprimer Loy482-NosT-upper-*Sal*I
5´-AAATTTGTCGACCGATCGGTCAAACATT-3´

20. SEQ ID NO: 56 Oligonucleotideprimer Loy483-NosT-Lower-*Hin*dIII
5´-AAATTTAAGCTTCCCGATCTAGTAACATAGATGACA-3´

21. SEQ ID NO: 57 Oligonucleotideprimer Loy494- Gus_upper_*Sal*I_Spacer
5´-TTTTAGTCGACACGCTGGACTGGCATGAACT-3´

22. SEQ ID NO: 58 Oligonucleotideprimer Loy492-NosT-lower- *Bgl*II *Spe*I
5´-TTTTAAGATCTACTAGTCCGATCTAGTAACATAGATGACA-3´

23. SEQ ID NO: 59 Oligonucleotideprimer Loy493_Gus_upper_*Sal*I_Spacer
5´- TTTAAGTCGACAAGTCGGCGGCTTTTCTGCT-3´

24. SEQ ID NO: 60 Oligonucleotideprimer Loy492-NosT-lower- *Bgl*II *Spe*I
5´-TTTTAAGATCTACTAGTCCGATCTAGTAACATAGATGACA-3´

25. SEQ ID NO: 61 Oligonucleotideprimer JMTOIprim1
5'-GGTTCCAAGGTACCAAAACAATGGGCGCTGATGATGTTGTT-GAT-3'

26. SEQ ID NO: 62 Oligonucleotideprimer JMTOIprim2
5'-AAGGTAGAAGCAGAAACTTACCTGGATACGTCACTTTGACCA-3'

27. SEQ ID NO: 63 Oligonucleotideprimer JMTOIprim3
5'-TGGTCAAAGTGACGTATCCAGGTAAGTTTCTGCTTCTACCTT-3'

28. SEQ ID NO: 64 Oligonucleotideprimer JMTOIprim4
5'-GGTTCCAAGGATCCATTTATTTTGAAAAAAATATTTG-3'

29. SEQ ID NO: 65 Oligonucleotideprimer JMTOIprim5
5'- GGTTCCAAGGATCCAGTATATAGCAATTGCTTTTC-3'

30. SEQ ID NO: 66 Oligonucleotideprimer JMTOIprim6
5'- CGAGAACCTTCGTCAGTCCTGCACATCAACAAATTTTGGTCAT-AAAAAAAAAAATATTAGAAAAGTTATAAATTAAAATATAC-3'

31. SEQ ID NO: 67 Oligonucleotideprimer JMTOIprim7
5'- CTAATATTTTTTTTTTTATGACCAAAATTTGTTGATGTGCAGGAC-TGACGAAGGTTCTCGCAC-3'

32. SEQ ID NO: 68 Oligonucleotideprimer JMTOIprim8
5'- TTGGAACCACTAGTTTATCGCCTGACACGATTTCCTGC-3'

33. SEQ ID NO: 69 Oligonucleotideprimer JMTOIprim9
5'- GGTTCCAAGGATCCGATCGTTCAAACATTTGGCAA-3'

34. SEQ ID NO: 70 Oligonucleotideprimer JMTOIprim10
5'- GGTTCCAAGGATCCGATCTAGTAACATAGATGACA-3'

35. SEQ ID NO: 71 Screening construct pJMTOI1

36. SEQ ID NO: 72 Screening construct pJMTOI2

37. SEQ ID NO: 73 Screening construct pJMTOI3

38. SEQ ID NO: 74 Screening construct pJMTOI4

39. SEQ ID NO: 75 Screening construct pJMTOI5

40. SEQ ID NO: 76 Lo376-pENTR-B2

41. SEQ ID NO: 77 Lo442 pSUN1-R4R3-M20 (OCS10) (destination vector)

42. SEQ ID NO: 78 Binary vector Lo239-pSUN3-GWs-B1-BnAK700::GUS::nosT-B2 (10414 bp)

43. SEQ ID NO: 79 Binary vector Lo657- pSUN3-GWs-B1-BnAK700::GUS::E9::nosT::B2 (11153 bp)

44. SEQ ID NO: 80 GFP-Primer5: 5'-CGGCCTAGGGGCGCCCGGACCGagctgttcaccggca-3'

45. SEQ ID NO: 81 GFP-Primer 6: 5'-CGG ACT AGT gat gta gcc ctc agg-3'

46. SEQ ID NO: 82 Primer 7: 5'- CGA GCT CGT GCC TTT TGG ATC G-3'

47. SEQ ID NO: 83 Primer 8: 5'- CGG TCC GAA CGT GGT TGG-3'

48. SEQ ID NO: 84 Primer 9: 5'- CGA GCT CGG CCC TAT GAA TTG G-3'

49. SEQ ID NO: 85 Primer 10: 5'- CGG TCC GTC TCC TTC TGC ACA C-3'

50. SEQ ID NO: 86 Primer 11: 5'-CGA GCT CGA TGC ATT CCT TGG AT-3'

51. SEQ ID NO: 87 Primer 12: 5'-CCT AGG GTT TGG AGG TAT CAA G-3'

52. SEQ ID NO: 88 Primer 13: 5'-CGA GCT CCG TCC GAT GTG ATT CCG TC-3'

53. SEQ ID NO: 89 Primer 14: 5'- CCT AGG GGC AGT GTC GGC GGT T-3'

54. SEQ ID NO: 90 Primer 15: 5'- CGA GCT CCA GAG TGA CAG ACA GTG A-3'

55. SEQ ID NO: 91 Primer 16: 5'- CCT AGG TCT TCA ACT GTC CCC A-3'

56. SEQ ID NO: 92 *Oryza sativa* terminator BPST.3 (1,137 bp). This sequence is a functional equivalent of the sequence described by SEQ ID NO: 45.

57. SEQ ID NO: 93 *Oryza sativa* terminator BPST.4 (reverse complementary sequence of BPST.3) (1,137 bp). This sequence is a functional equivalent of the sequence described by SEQ ID NO: 45

58. SEQ ID NO: 94 Artificial sequence, vector pRJB058 (6,849 bp))

59. SEQ ID NO: 95 insert from pRJB062: Nos terminator (Nos-T) sequence inserted into *Sac*I-*Rsr*II fragment of pRJB058 (257 bp)

60. SEQ ID NO: 96 insert from pRJB064: ORF sequence inserted into *Sac*I digested and T4 DNA Polymerase filled in fragment of pRJB058 (1,089 bp)

61. SEQ ID NO: 97 insert from pRJB066: *Oryza sativa* BPST.1 sequence inserted into *Sac*I digested and T4 DNA polymerase filled in fragment of pRJB058 (1,420 bp)

62. SEQ ID NO: 98 insert from pRJB065: *Oryza sativa* BPST.2 sequence inserted into *Sac*I digested and T4 DNA polymerase filled in fragment of pRJB058 (1,414 bp)

63. SEQ ID NO: 99 insert from pRJB067: *Oryza sativa* BPST.3 sequence inserted into *Sac*I digested and T4 DNA polymerase filled in fragment of pRJB058 (1,165 bp)

64. SEQ ID NO: 100 insert from pRJB068: *Oryza sativa* BPST.4 (reverse complementary sequence of BPST.3) sequence inserted into *Sac*I digested and T4 DNA polymerase filled in fragment of pRJB058 (1,165 bp)

65. SEQ ID NO: 101 BPST.5-MCS: *Oryza sativa* BPST.5 sequence with *Eco*RI and *Avr*II sites (1,305 bp)

66. SEQ ID NO: 102 BPST.6-MCS: *Oryza sativa* BPST.6 sequence with *Eco*RI and AvrII sites (1,350 bp)

67. SEQ ID NO: 103 BPST.7-MCS: *Oryza sativa* BPST.7 sequence with *Eco*RI and *Sac*I sites (1,532 bp)

68. SEQ ID NO: 104 BPST.8-MCS: *Oryza sativa* BPST.8 sequence (reverse complementary sequence of BPST.7) with *Eco*RI site (1,532 bp)

69. SEQ ID NO: 105 binary vector pRLI024 derived from pRJB058 (16,914 bp)

70. SEQ ID NO: 106 binary vector pRLI031 derived from pRLI024 (15,919 bp)

71. SEQ ID NO: 107 *Oryza sativa* terminator BPST.7 (1,499 bp). This sequence is a functional equivalent of the sequence described by SEQ ID NO: 11.

72. SEQ ID NO: 108 *Oryza sativa* terminator BPST.8 (reverse complementary sequence of BPST.3) (1,499 bp). This sequence is a functional equivalent of the sequence described by SEQ ID NO: 11.

## Examples

### Chemicals

Unless indicated otherwise, chemicals and reagents in the Examples were obtained from Sigma Chemical Company (St. Louis, MO), restriction endonucleases were from New England Biolabs (Beverly, MA) or Roche (Indianapolis, IN), oligonucleotides were synthesized by MWG Biotech Inc. (High Point, NC), and other modifying enzymes or kits regarding biochemicals and molecular biological assays were from Clontech (Palo Alto, CA), Pharmacia Biotech (Piscataway, NJ), Promega Corporation (Madison, WI), or Stratagene (La Jolla, CA). Materials for cell culture media were obtained from Gibco/BRL (Gaithersburg, MD) or DIFCO (Detroit, MI). The cloning steps carried out for the purposes of the present invention, such as, for example, restriction cleavages, agarose gel electrophoresis, purification of DNA fragments, transfer of nucleic acids to nitrocellulose and nylon membranes, linking DNA fragments, transformation of *E. coli* cells, growing bacteria, multiplying phages and sequence analysis of recombinant DNA, are carried out as described by Sambrook (1989). The sequencing of recombinant DNA molecules is carried out using ABI laser fluorescence DNA sequencer following the method of Sanger (Sanger 1977).

### Example 1: Development of genomic libraries for identification of transcription terminators

Genomic DNA from a target plant is prepared according to Qiagen plant DNA preparation kit (cat# 12143). One μg of the genomic DNA is digested with four base pair cutting enzyme (*e.g. Sau*3A) overnight at 37°C or mechanical shearing in a Hamilton syringe or sonication followed by electroporation (0.8% Agarose gel) and gel purification using the QIAEX II Gel Extraction Kit (cat# 20021). Fragmented genomic DNAs (500 to 1,000 bp) are cloned in to the screening constructs or screening vectors described herein. The resulting library of constructs and vectors is batch transformed into plant cells (see below).

### Example 2: *In silico* identification of sequences to evaluation as transcription terminators

Beside other approaches described herein to provide sequences for evaluation of their suitability as transcription terminator sequences (*e.g.*, genomic sequences as provided by Example 1), sequences can be provided by *in silico* search of genome databases, such as for example of *Arabidopsis thaliana* or rice. Accordingly, the whole rice genome sequences are screened for potential plant derived terminator candidates using the most updated data from the Institute of Genome Research (TIGR; PUB_tigr_rice_genome_v4.nt (v03212003), PUB_tigr_rice_cds_Oct022003.nt, pubOSest0603 (ncbi)). This screening system comprises three major components:

1) identification of paired genes meeting predefined (for example 700 to 2,000 bp) intergenic distance criteria;
2) determination of the expression levels and expression patterns of the identified paired genes;
3) selection of intergenic sequences for terminator candidates by genome mapping.

The genome mapping requires the following activities: (1) manual verification of the gene model, reading frame of the coding sequences (CDS), and the intergenic struc-

tures underlying the genomic sequences and (2) selection of potential transcription terminators of interest candidates based on the EST sequence alignment and CDS.

**2.1 Identification of paired genes of interest**

Given the recently released rice japonica genome sequences and the 56,056 annotated rice CDS, the coordinators of the beginning and ending of those 56,056 annotated rice CDS from genomic regions are retrieved and the intergenic distances are calculated. A frequency distribution table of intergenic distances at 200 bp/interval is generated such that appropriate intergenic distance can be defined. In order to capture maximal values of potential terminator candidates, the distance between genes in the range of 700 to 2,000 bp is used, leading to identify 16,058 pairs of rice genes (consisting of paired genes in head-to-tail, tail-to-head, and tail-to-tail orientation).

First, each pair of the identified rice CDS is blasted against rice EST databases to retrieve EST homolog sequences. The identified sequences that are homologous to ESTs are mapped onto the same rice genomic regions from which the rice CDS are derived using the splice alignment gene identification application, GeneSeqer™ (Version 1.9 (October 22, 2002), Department of Zoology & Genetics, Iowa State University, Ames, IA 50011-3260). The underlying gene model, including the 5' end exon, the 3' end exon, CDS reading frame, and intergenic structure between two genes is carefully verified by graphically displaying the GeneSeqer™ genome mapping results using MyGV (Version 1.0 (from NewLink Genetics, 2901 South Loop Drive, Suite 3900, Ames, Iowa 50010) application. Potential gene terminator candidates are

1) the paired CDS reading frames must be either head-to-tail, tail-to-tail, or tail-to-head orientated. The tail-to-tail orientation (*i.e.* from which expression from said genes is directed in opposite direction against each other) is the most desirable, as the intergenic sequences do not contain the promoter sequences and the intergenic sequence length can be minimized;
2) the annotated CDS and its gene model must be verified and supported by the EST sequences according to sequence alignment.

Of these intergenic regions preferably regions are identified for further analysis which are localized between genes which have a tail-to-tail localization (*i.e.* from which expression from said genes is directed in opposite direction against each other).

**2.2 Determination of gene expression levels and expression patterns**

Each pair of the identified rice CDS (*i.e.* corresponding to the genes flanking the intergenic region) is used to identify the corresponding EST sequences of high identity to rice EST database using blastn searching with expectation value set to $1.0e^{-20}$. Those identified EST sequences, which presumably are considered as the same sequences as the rice CDS, are used to retrieve the gene expression profiling data derived from either the cDNA library clone distribution or microarray expression. Overall, a gene with a cluster/variant size of more than 100 clones derived from the cDNA libraries is considered as highly expressed, and so does the signal intensity beyond the top 25% quantile from the microarray expression studies. Highly expressed abundance for both of the paired genes is required as criteria for gene selection. Furthermore, the co-expression pattern of the paired genes can be assessed using the clone distribution across cDNA libraries or using the microarray expression data across different experiments. A linear correlation coefficient is calculated to determine the pattern of the gene

expression. A pair of genes demonstrating unique expression pattern is desirable. Using those criteria, 5,279 pairs of rice CDs sequences are selected.

### 2.3 Determination of Transcript Length Variability

Preferably, the 3'-end of the EST sequence alignments corresponding to the genes flanking the intergenic region must demonstrate a low degree of variability with respect to transcript length. This is found to be predictive for a strong terminator signal.

Based on the above criteria, 37 rice potential intergenic genomic sequences (SEQ ID NO: 10 to 46) are selected for testing in the screening systems of the invention in order to identify terminators of interest. All of these sequences are localized in between genes which are orientated in the above mentioned preferred tail-to-tail orientation.

## Example 3: *In vitro* screening system for identifying terminators of interest

A high throughput screening method is developed to identify transcription terminators at the mRNA levels. The method includes *in vitro* transcription using single strand fluorescence probes such as beacon probes that hybridize polyadenylated RNA region and the read through region. The fluorescence amount of the read through products are compared with the amount of polyadenylated RNA. The stronger and tighter terminators will show the lesser amounts of read through products. Control vectors are constructed to establish the screening system (see Fig. 13 and agenda to this figure above).

A promoter for these constructs is preferably a strong constitutive promoter (*e.g.* maize ubiqutin promoter). In order to measure uncoupled transcription, SP6 or T7 phase promoter can be used for *in vitro* transcription. The coding sequence in the expression cassette can be any reporter gene or genomic DNA including start at the 5' end and stop codon at the 3' end, which do not have sequence homology to plant genome (*e.g.* intergenic sequences from yeast genome). Nopaline synthase terminator can be replaced with any other known terminator to use as a control or uncharacterized genomic DNA fragment to identify potential terminator candidates.

### 3.1 Vector Construction

Vector pBPSMM268 contains the GUS::potato intron gene followed by the NOS terminator region. To this vector, maize Ubiquitin promoter::intron is added by digestion of pMM268 with *Stu*I and *Sma*I, followed by blunt ligation of the Ubiquitin promoter::intron fragment obtained from *Stu*I digestion of pBPSCER043, which produces vector pTOI01.

In order to ensure efficient transcript processing of mRNAs that do not undergo transcriptional termination at putative transcription terminators, the OCS terminator region is cloned into pTOI01. pTOI02 is generated by digestion of pTOI01 with *Sac*I and *Not*I, and ligation of the *Sac*I/*Not*I fragment generated from the PCR amplification of the OCS terminator from vector p1bxSuperGusQC using primers 1 and 2 (SEQ ID NO 47 and 48).

Primer 1 (*Sac*I, *Avr*II, *Spe*I, OCS 5'; SEQ ID NO: 47):
5'-CG GAGCTC CCTAGG ACTAGT tcgaccggcatgccc-3'

Primer 2 (*Not*I, OCS 3'; SEQ ID NO: 48):
5'- CC GCGGCCGC agcttggacaatcag-3'

A fragment of the firefly luciferase gene is cloned downstream of the transcription terminator sequences to be assessed insertion site in order to act as a unique sequence that is only transcribed in the presence of a poorly functioning terminator. pTOI03 is generated from the digestion of pTOI02 with *Avr*II and *Spe*I, and ligating in the *Avr*II/*Spe*I fragment generated from the PCR amplification of a 240bp fragment of the firefly luciferase gene ($Lu^F$ ) from vector pGL3 (R2.2) basic vector (Promega cat# E6441) using primers 3 and 4 (SEQ ID NO: 49 and 50).

Primer 3 (*Avr*II, *Xma*I, *Rsr*II, $Lu^F$ 5'; SEQ ID NO: 49):
5'-CG CCTAGG CCCGGG CGGACCG cattaagaagggccc-3'

Primer 4 (*Spe*I $Lu^F$ 3'; SEQ ID NO: 50):
5'-CG ACTAGT agagagttctcagagc-3'

Vector pTOI03 is the base vector that is used to generate constructs testing putative transcription terminator sequences. Vector pTOI04 comprises pTOI03 with the addition of the NOS in forward orientation, and is generated by insertion of the NOS containing *Sac*I fragment from pBPSCR043 into the unique *Sac*I site of pTOI03. (Positive control – NOS). Vector pTOI05 comprises pTOI03 with the addition of the NOS in reverse orientation, and is generated by the insertion of the inverted NOS *Sac*I fragment from pBPSCR043 into the unique *Sac*I site of pTOI03. (Negative control – inverted NOS).

Vectors pTOI06 – pTOI10 are generated by the PCR amplification of putative terminator sequences from rice genomic DNA (selected from the seuences decribed by SEQ ID NO: 10 to 46) such that a *Sac*I site is generated on the 5' end of the sequence and a *Rsr*II site is generated on the 3' end. (Note: if the sequence of individual genomic elements precludes the use of these two restriction enzymes, then the alternative enzymes *Avr*II or *Xma*I can be used for cloning purposes.) The source of transcription terminators can be from both the *in silico* screening system and the genomic libraries containing 500 to 1,000 bp fragments.

**3.2 Preparation of BMS suspension cultures cells**

Black Mexican Sweetcorn (BMS) suspension cultured cells are propagated in Murashige and Skoog (MS) liquid medium containing 2% (w/v) sucrose and 2 mg/L 2,4-dichlorophenoxyacetic acid. Every week 5 mL of a culture of stationary cells are transferred to 125 mL of fresh medium and cultured on a rotary shaker operated at 130 rpm at 27°C in a 500 mL flask in the dark.

**3.3 Preparation of the nuclear extract**

The HeLa nuclear extract are purchased from Promega (HeLaScribe® Nuclear Extract; cat# E3092). Nuclear extracts are prepared from BMS cells as described (Moreno *et al.*, 1997). BMS suspension cultured cells at logarithmic phase are harvested three days after the start of a fresh culture by spinning down at 2 krpm for 500 mL tubes for 10 min at 4°C (1,200 rpm for 800 mL glass conical bottles at 170 x g). The cell pellet is loosened and resuspended in cold HBSS (Hank's Balanced Salt Solution; Sigma cat# H9269). The cells are transferred into 50 mL Corning tube and spanned down at 1,200

rpm at 4°C. Packed cell volume (PCV) is measured by eye. The pellet is loosened and resuspended in 5x PCV hypotonic buffer followed by swelling the cells on ice for 10 min. The cells are spanned down at 1,200 rpm for 10 min. The supernatants are removed. One volume of PCV hypotonic buffer including 0.1% NP-40 is added to the pellet followed by resuspending the cells. The resuspended cells are transferred into chilled dounce homogenizer and measured the total volume before adding 1x PMSF (500x: 8.71 mg/mL). The cells are dounced for 10 to 15 strokes and checked the cells to yield 80 to 90% cell lysis. It is critical to avoid overdouncing the cells. Trypan blue is added to a small portion of the cells to check cell lysis under microscope. Blue cells indicate cell lysis. The cell lysis is quickly transferred into Corex 30 mL tube. 0.1 volume of sucrose restore buffer is added and gently mixed. The rotor and centrifuge have to be pre-cold. The nuclei are immediately spanned down at 10 krpm for 2 min in Beckman JA-20 rotor with brake. The supernatants containing cytoplasm are carefully removed and saved by adding glycerol to 20% (v/v) and stored at -70°C. The pellet is detached using a pipette and transferred into the nuclear resuspension buffer (3 mL/$10^9$ cells) in an ultracentrifuge tube followed by adding N-α-tosyl-L-lysine chloromethyl ketone (TLCK) protease inhibitor (250x: 10 mg/mL in 1 mM HCl), leupeptin (2,000x: 1 mg/mL in $dH_2O$), aprotinin (Sigma cat# A1153; 1,000x: 1 mg/mL in $dH_2O$), and pepstatin A (Sigma cat# P4265; 2,000x: 1 mg/mL in MeOH) to 1x. The tubes are balanced, rocked gently for 30 min, and spanned at 35 krpm in Ti454 (or 42 krpm in Ti70.a) for 90 min at 2°C (150,000 x g). The supernatants are transferred into another ultracentrifuge tube and measured the volume by eye. 0.33g $(NH_4)_2SO_4$/mL is sprinkled into the extract for over 30 min with stirring or rocking until salt is dissolved after each addition on ice. The solution turns milky as the protein precipitates and is stirred or rocked for an additional 20 min at 4°C followed by spinning down at 35 krpm for 30 min in Ti45 (or 32 krpm in Ti70.1). The pellet is resuspended in less than 1 mL of dialysis buffer ($10^9$ cells/mL). The resuspended cells are dialyzed for one hour against more than 200 volume of dialysis buffer (2 L). The buffer should be changed during dialysis for an additional four hours. The dialyzed extract is spanned down at 35 krpm for one hour followed by storing small aliquots at –80°C.

<u>Hypotonic Buffer</u>
10 mM HEPES, pH 7.9 (KOH)
0.75 mM spermidine
0.15 mM spermine
0.1 mM EDTA
0.1 mM EGTA, pH7.5 (KOH)
1 mM DTT
10 mM KCl
(add protease inhibitors and DTT before use)

<u>10x Sucrose Restore Buffer</u>
500 mM HEPES, pH 7.9 (KOH)
7.5 mM spermidine
1.5 mM spermine
10 mM KCl
2 mM EDTA
10 mM DTT

1 x Sucrose Restore Buffer = 1 volume 10 x salts + 9 volume 75% (w/v) sucrose

Nuclear Resuspension Buffer
20 mM HEPES, pH 7.9 (KOH)
0.75 mM spermidine
0.15 mM sermine
0.2 mM EDTA
2 mM EGTA, pH 7.5 (KOH)
1 mM DTT
25% glycerol
10% saturated ammonium sulfate
(add protease inhibitors and DTT before use)

Dialysis Buffer
20 mM HEPES, pH7.9 (KOH)
20% glycerol
100 mM KCl
0.2 mM EDTA
0.2 mM EGTA, pH 7.5 (KOH)
2 mM DTT
(add protease inhibitors and DTT before use)

**3.4 An *in vitro* assay system**

Primer sequences for molecular beacon probes are chosen (1) between GUS and NOS for detecting polyadenylated products and (2) within the truncated firefly luciferase gene for detecting read through products. The probes are designed by using Beacon Designer 3.0. Two different reporter dyes are chosen for this assay (*e.g.* Texas Red, Rhodamine Red, Tamra, Joe, Tox, Oregon green, *etc.*).

The constructs are linearized by restriction enzyme digestion with *Not*I enzyme at 37°C overnight followed by electroporation (0.8% Agarose gel) and gel purification using the QIAEX II Gel Extraction Kit (cat# 20021). One μg of the linearized single template is added into the reaction solution in a total volume of 25 μL (15 μL of a mixture of HeLa and BMS nuclear extracts at 1:1 ratio [v/v], 400 μM ATP, CTP, GTP, UTP, 400 nM final concentration of two beacon probes, 5mM $MgCl_2$, mg/mL BSA). The reaction solution is incubated for 2 hour at room temperature. The reaction progress is monitored using a Cytofluor multiwell plate reader at an excitation wavelength of 485 nm and an emission wavelength of 530 nm. [Razik and Quatrano, 1997; Yammaguchi *et al.*, 1998; Liu *et al.*, 2002]. If a particular transcription terminator sequences to be assessed provides efficient transcriptional termination, the expression of sequences complimentary to probe 1 is much greater than the expression of probe 2-specific sequences. If a sequence does not terminate efficiently the ratio of probe 1: probe 2 expression is lower. A ratio of the yield obtained between polyadenylated RNA and the read through products is calculated to determine potential terminator candidates (see Fig. 13B and agenda to this figure above). In addition to using single strand fluorescence probes, the ratio of the yield can be detected using Reverse Transcriptase (RT)-PCR following the protocols in the art.

## Example 4: *In vivo* screening system for identifying terminators of interest

### 4.1 Vector Construction

4.1.1 pUC expression vectors

Vector pBPSMM268 contains the GUS::potato intron gene followed by the NOS terminator region. To this vector, maize Ubiquitin promoter::intron was added by digestion of pMM268 with *Stu*I and *Sma*I, followed by blunt ligation of the Ubiquitin promoter::intron fragment obtained from *Stu*I digestion of pBPSCER043, which produced vector pRJB051.

A fragment of the green fluoresecent protein (GFP-f) gene was cloned downstream of the transcription terminator sequences to be assessed insertion site in order to act as a unique sequence that is only transcribed in the presence of a poorly functioning terminator. pRJB058 was generated from the digestion of pRJB051 with *Avr*II and *Spe*I, and ligation of the *Avr*II/*Spe*I fragment generated from the PCR amplification of a 260bp fragment of the GFP gene (GFP-F ) from vector pALGFP1 using GFP-primers 5 and 6 (SEQ ID NOs: 80 and 81).

GFP-Primer 5 (*Avr*II, *Kas*I, *Rsr*II, AKR GFP19 5'; SEQ ID NO: 80):
5'-CGG CCT AGG GGC GCC CGG ACC Gag ctg ttc acc ggc a-3'

GFP-Primer 6 (*Spe*I, S GFP 281 3'; SEQ ID NO: 81):
5'-CGG ACT AGT gat gta gcc ctc agg-3'

Vector pRJB058 is the base vector that was used to generate constructs testing putative transcription terminator sequences. Vector pRJB062 (SEQ ID NO: 95) comprises pRJB058 (SEQ ID NO: 94) with the addition of the NOS in forward orientation, and was generated by insertion of the NOS containing *Sac*I fragment from pBPSCR043 into the unique *Sac*I site of pTOI03. (Positive control –NOS).

Vector pRJB063 (reverse complementary sequence of SEQ ID NO: 95) comprises pRJB058 with the addition of the NOS in reverse orientation, and is generated by the insertion of the inverted NOS *Sac*I fragment from pBPSCR043 into the unique *Sac*I site of pTOI03. (Negative control – inverted NOS).

Vector pRJB064 (SEQ ID NO: 96) comprises pRJB048 with the addition of the 1.1Kb ORF fragment from pRJB018. This vector will serve as a negative control for specific transcriptional termination by putative TOIs, as the sequence comprises an internal fragment from a known open reading frame, and should therefore possess minimal intrinsic transcriptional termination activity. Vector pRJB064 was generated by ligation of the 1.1Kb *Hpa*I/*Stu*I fragment from pRJB018 into *Sac*I digested and 3'-5- exonuclease-treated pRJB058.

Vectors pRJB065 (SEQ ID NO: 98) and pRJB066 (SEQ ID NO: 97) comprise pRJB058 with the addition of the rice genomic DNA BPST.2 (reverse complementary sequence of SEQ ID NO:33) and BPST.1 (SEQ ID NO:33), respectively. The 1.4Kb PCR product produced from amplification of rice genomic DNA with primers 7 and 8 (SEQ ID Nos 82 and 83):

Primer 7: 5'- CGA GCT CGT GCC TTT TGG ATC G-3'
Primer 8: 5'- CGG TCC GAA CGT GGT TGG-3'

The PCR product was TOPO cloned to produce pTOPO BPST.1 (SEQ ID NO:33) and BPST.2 (reverse complementary sequence of SEQ ID NO:33). The 1.4 Kb fragment resulting from *Eco*RI digestion and T4 DNA polymerase fill in reaction of pTOPO BPST.1 and BPST.2 was ligated into *Sac*I digested and 3'-5- exonuclease-treated pRJB058. pRJB065 (SEQ ID NO: 98) represents the resulting vector comprising the BPST.2 putative terminator, and pRJB066 (SEQ ID NO: 97) represents the ligation product comprising the BPST.1 sequence.

Vectors pRJB067 (SEQ ID NO: 99) and pRJB068 (SEQ ID NO: 100) comprise pRJB058 with the addition of the rice genomic DNA BPST.3 (SEQ ID NO:92) and BPST.4 (reverse complementary sequence, SEQ ID NO:92), respectively. The 1.1Kb PCR product produced from amplification of rice genomic DNA with primers 9 and 10 (SEQ ID NO: 84 and 85):

Primer 9: 5'- CGA GCT CGG CCC TAT GAA TTG G-3'
Primer 10: 5'- CGG TCC GTC TCC TTC TGC ACA C-3'

The PCR product was TOPO cloned to produce pTOPO BPST.3 and BPST.4. The 1.1 Kb fragment resulting from *Eco*RI digestion and T4 DNA polymerase fill in reaction of pTOPO BPST.3 and BPST.4 was ligated into *Sac*I digested and 3'-5-exonuclease-treated pRJB058. pRJB067 (SEQ ID NO: 99) represents the resulting vector comprising the BPST.3 putative terminator, and pRJB068 (SEQ ID NO: 100) represents the ligation product comprising the BPST.4 sequence.

BPST.5 (SEQ ID NO:18) produced a 1.2 Kb PCR product from amplification of rice genomic DNA with primers 11 and 12 (SEQ ID NO: 86 and 87):

Primer11: 5'-CGA GCT CGA TGC ATT CCT TGG AT-3'
Primer12: 5'-CCT AGG GTT TGG AGG TAT CAA G-3'

BPST.6 (SEQ ID NO:10) produced a 1.3Kb PCR product from amplification of rice genomic DNA with primers 13 and 14 (SEQ ID NO:88 and 89):

Primer 13: 5'-CGA GCT CCG TCC GAT GTG ATT CCG TC-3'
Primer 14: 5'- CCT AGG GGC AGT GTC GGC GGT T-3'

BPST.7 (SEQ ID NO:107) and BPST.8 (reverse complementary sequence of SEQ ID NO:108) produced a1.5 Kb PCR product from amplification of rice genomic DNA with primers 15 and 16 (SEQ ID NO:90 and 91):

Primer 15: 5'- CGA GCT CCA GAG TGA CAG ACA GTG A-3'
Primer 16: 5'- CCT AGG TCT TCA ACT GTC CCC A-3'

Additional TOI candidates will be isolated and cloned into pUC expression vectors as described above.

**4. 1. 2 Binary vectors**

For evaluation of transcriptional termination by the putative TOI sequences in stably transformed maize plants, binary vectors were prepared for *Agrobacterium*-mediated maize transformation. The full-length T-DNA sequences for vectors pRLI024 and pRLI031 are provided in the attached sequence listing (SEQ ID NO: 105 and 106, respectively). The other vectors were derived therefrom by exchanging the terminator regions.

Vector pRLI024 (SEQ ID NO:105) was generated by ligation of the 4.9Kb *Pvu*II fragment from pRJB058 into pLM150 that had been digested with *Pme*I, generating pJB077. The 3.1Kb DsRed2 expression cassette was liberated from vector pLM299 via *Fse*I/*Pac*I digestion, and ligated into *Fse*I/*Pac*I digested pRJB077 to generate pRLI024 (SEQ ID NO:105).

Vector pRLI025 was generated by ligation of the 4.9Kb *Pvu*II fragment from pRJB062 (SEQ ID NO:95) into pLM150 that had been digested with *Pme*I, generating pJB078. The 3.1Kb DsRed2 expression cassette was liberated from vector pLM299 via *Fse*I/*Pac*I digestion, and ligated into *Fse*I/*Pac*I digested pRJB078 to generate pRLI025.

Vector pRLI026 was generated by ligation of the 4.9Kb *Pvu*II fragment from pRJB064 (SEQ ID NO:96) into pLM150 that had been digested with *Pme*I, generating pJB079. The 3.1Kb DsRed expression cassette was liberated from vector pLM299 via *Fse*I/*Pac*I digestion, and ligated into *Fse*I/*Pac*I digested pRJB079 to generate pRLI026.

Vector pRLI027 was generated by ligation of the 4.9Kb *Pvu*II fragment from pRJB066 (SEQ ID NO: 97) into pLM150 that had been digested with *Pme*I, generating pJB080. The 3.1Kb DsRed expression cassette was liberated from vector pLM299 via *Fse*I/*Pac*I digestion, and ligated into *Fse*I/*Pac*I digested pRJB080 to generate pRLI027.

Vector pRLI028 was generated by ligation of the 4.9Kb *Pvu*II fragment from pRJB065 (SEQ ID NO: 98) into pLM150 that had been digested with *Pme*I, generating pJB081. The 3.1Kb DsRed expression cassette was liberated from vector pLM299 via *Fse*I/*Pac*I digestion, and ligated into *Fse*I/*Pac*I digested pRJB081 to generate pRLI028.

Vector pRLI029 was generated by ligation of the 4.9Kb *Pvu*II fragment from pRJB067 (SEQ ID NO:99) into pLM150 that had been digested with *Pme*I, generating pRLI022. The 3.1Kb DsRed expression cassette was liberated from vector pLM299 via *Fse*I/*Pac*I digestion, and ligated into *Fse*I/*Pac*I digested pRLI022 to generate pRLI029.

Vector pRLI030 was generated by ligation of the 4.9Kb *Pvu*II fragment from pRJB068 (SEQ ID NO:100) into pLM150 that had been digested with *Pme*I, generating pRLI023. The 3.1Kb DsRed expression cassette was liberated from vector pLM299 via *Fse*I/*Pac*I digestion, and ligated into *Fse*I/*Pac*I digested pRLI023 to generate pRLI030.

An alternative series of binary vectors was generated in order to evaluate putative TOIs with regard to their ability to direct bi-directional transcriptional termination. For these vectors, the TOI sequences were cloned between two reporter expression cassettes in tail-to-tail orientation.

Vector pLI024 (SEQ ID NO: 105) was digested with *Sac*I to remove the 950bp Nos-T and intervening sequences from between the DsRed ORF and the TOI MCS. The vector was recircularized to generate pLI031 (SEQ ID NO:106).

Vector pLI025 was digested with *Sac*I to remove the 950bp Nos-T and intervening sequences from between the DsRed ORF and the TOI MCS. The vector was recircularized to generate pLI032.

Vector pLI026 was digested with *Sac*I to remove the 950bp Nos-T and intervening sequences from between the DsRed ORF and the TOI MCS. The vector was recircularized to generate pLI033.

Vector pLI027 was digested with *Sac*I to remove the 950bp Nos-T and intervening sequences from between the DsRed ORF and the TOI MCS. The vector was recircularized to generate pLI034.

Vector pLI028 was digested with *Sac*I to remove the 950bp Nos-T and intervening sequences from between the DsRed ORF and the TOI MCS. The vector was recircularized to generate pLI035.

.

Vector pLI029 was digested with *Sac*I to remove the 950bp Nos-T and intervening sequences from between the DsRed ORF and the TOI MCS. The vector was recircularized to generate pLI036.

Vector pLI030 was digested with *Sac*I to remove the 950bp Nos-T and intervening sequences from between the DsRed ORF and the TOI MCS. The vector was recircularized to generate pLI037.

Additional TOI candidates will be isolated and cloned into binary vectors as described above.

**4.2 Assays for identifying terminators of interest**

The test construct comprised a GUS reporter gene driven by the maize ubiquitin promoter, and enzyme sites to insert putative TOI and control sequences. The TOI functionality screen was based on the principle that in the absence of a functional terminator region the GUS mRNA will not be efficiently processed, and therefore will not be available to support high levels of translation of GUS protein. The results of these transient analyses are shown in Table 2 and Figure 14. The experimental rationale was supported by the finding that the vector lacking an insertion at the TOI cloning site does not drive detectable GUS expression (pRJB058; SEQ ID NO: 94). The insertion of the nopaline synthase (Nos) terminator was able to rescue GUS expression (pRJB062; SEQ ID NO: 95). The insertion of sequence derived from an internal portion of an exogenous protein-coding gene (ORF sequence) did not result in GUS expression (pRJB064; SEQ ID NO: 96), signifying that the GUS expression seen with pRJB062 was due to intrinsic transcriptional termination activity by Nos, and not a non-specific effect due to insertion of any DNA sequence at that site. BPST.1 (SEQ ID NO:33) and BPST.2 (reverse complementary sequence of SEQ ID NO:33) showed GUS expression levels comparable to that seen with the Nos (+) control vector. BPST..3 (SEQ ID NO:92) and BPST.4 (reverse complementary sequence, SEQ ID

NO:92) consistently resulted in significant GUS expression at levels that appeared to be slightly lower than observed with Nos.

Table 2. Transient GUS expression testing for terminator candidates

| Terminator candidates | GUS expression* | |
|---|---|---|
| BPST (null) - (pRJB058; SEQ ID NO: 94) | - | 0% |
| BPST (+) - Nos (pRJB062; SEQ ID NO: 95) | ++++ | 100% |
| BPST (-) - ORF (pRJB064; SEQ ID NO: 96) | - | 0% |
| BPST.1 (pRJB066; SEQ ID NO: 97) | ++++ | 100% |
| BPST.2 (pRJB065; SEQ ID NO: 98) | ++++ | 100% |
| BPST.3 (pRJB067; SEQ ID NO: 99) | +++ | 80% |
| BPST.4 (pRJB068; SEQ ID NO: 100) | +++ | 80% |

*GUS histochemical assays: a range of GUS activities (- no expression to ++++ high expression).

No GUS staining was observed in vectors that do not comprise a functional transcriptional terminator downstream of the GUS coding sequence (pRJB058 and pRJB064; SEQ ID NO: 94 and 96, respectively). The presence of a functional terminator rescued GUS expression in the (+) control (pRJB062; SEQ ID NO: 95) vector as well as all four TOI candidate sequences (pRJB065-pRJB068: SEQ ID indicated in Table 2) (Figure 14).

**4.3 Analysis of terminator candidates in stably transformed maize**

The binary vectors pBPSLI027, pBPSLI028, pBPSLI029, and pBPSLI030 will be transformed into maize using *Agrobacterium*-mediated transformation (Example 11.4). The levels and patterns of GUS expression controlled by BPST.1 (SEQ ID NO:33), BPST.2 (reverse complementary sequence of SEQ ID NO:33)., BPST.3 (SEQ ID NO: 92), or BPST.4 (reverse complementary sequence of SEQ ID NO:92) terminator will be compared with those controlled by NOS-t. BPST.1, BPST.2, BPST.3 and BPST.4 should show similar levels to that observed in transient assays (Table 2). This result will indicate that a transient assay can be used as a model system and is therefore one of the important screening systems to identify functional transcriptional terminators. However, the results obtained with the transient assays should be validated by the production of stable transformed transgenic plants.

**Example 5: *In vivo* screening system using gene silencing for identifying potential transcription terminators**

A high throughput screening system is developed to identify and isolate tight transcriptional termination sequences. This method is time-efficient and does not involve RNA analysis. Since dsRNA molecules are efficient in even a small amount, only very tight terminators can be identified. (Fig. 13C)

As described above in more detail, the RNA may be preferably selected from

a) RNAs encoding for an antisense or preferably double-stranded RNAs, which down-regulates expression of essential plant genes. In principle, any gene is suitable which has a lethal phenotype in a homozygous knockout (*e.g.* Phytoen desaturase, Nitrate reductase, HPPD, Acetohydroxyacid synthase, *etc.*)

b) RNAs encoding for toxic proteins, which expression causes lethal effect to the transgenic plants (negative selection markers like *e.g.*, *TK, codA, tyrA*, Diphtheria toxin *etc.*). Furthermore, any endogenous gene suitable as herbicidal target can be employed in the above-mentioned approach (Table 1).

**5.1 Vector construction for RNAi**

For experiments using RNAi down-regulation mechanism to evaluate transcription termination, vector pTOI03 is modified such that $Lu^F$ is replaced with an appropriate dsRNA cassette. An appropriate RNAi cassette comprises a 200 to 300 bp sequence that is specific to the targeted gene, followed by a spacer sequence of approximately 150-200bp, followed by an inverted repeat of the gene specific sequence, such when transcribed, a hairpin structure is formed with the gene-specific sequence forming the dsRNA stem. The RNAi target can be any gene that provides lethality or a screenable phenotype under down-regulation (*e.g.* AHAS, bar, target genes of herbicides (see Table 1), or other essential endogenous genes such as housekeeping genes). Gene-specific sequences are generated via PCR with appropriate restriction sites for forward and reverse orientation via amplification with primers 5 and 6 (SEQ ID NO: 51 and 52):

Primer 5 (*Rsr*II, *Bsp*EI, target gene seq 5'; SEQ ID NO: 51):
5' CG CGGACCG TCCGGA-N-3'
[N: gene-specific sequence of preferably 10 to 20 bases]

Primer 6 (*Spe*I, *Age*I, target gene seq 3'; SEQ ID NO: 52):
5'-CG ACTAGT ACCGGT-N-3'
[N: gene-specific sequence of preferably 10 to 20 bases]

The RNAi vector (pTOI11) is produced via a four-way ligation between (1) pTOI03 digested with *Rsr*II and *Spe*I, (2) target gene PCR product digested with *Rsr*II and *Age*I, (3) the spacer sequence with *Age*I and *Bsp*EI ends, and (4) target gene PCR product digested with *Bsp*EI and *Spe*I.

**5.2 Assays for identifying terminators of interest**

These experiments are performed by bombardment of plant tissues or culture cells (Example 9.1), by PEG-mediated (or similar methodology) introduction of DNA to plant protoplasts (Example 9.2), or by *Agrobacterium*-mediated transformation (Example 9.3). The target tissues for these experiments can be plant tissues (*e.g.* leaf or root), cultured cells (*e.g.* maize BMS), or plant tissues (*e.g.* immature embryos) for *Agrobacterium* protocols.

The sequences used as potential transcription terminator sequences can either be derived from the *in silico* transcription terminator sequence screen or from a library of random genomic fragments. Only plants can survive in subsequent regeneration in the case of stable transformation, which have an efficient terminator inserted in front of the sequence encoding the toxic RNA (thereby blocking its expression). The surviving plants are isolated and the terminator sequence amplified using PCR and sequencing.

**Example 6: *In vivo* screening system using bicistronic RNA detection**

A system is developed that utilizes the internal ribosome entry site (IRES) from encephalomiocarditis virus (EMCV), a picornovirus, or any functional IRES in plants. EMCV IRES has been shown in plants to efficiently direct translation of internally encoded proteins in parallel with canonical cap-mediated translation (Urwin *et al.*, 2000). This method allows the screening of potential terminator sequences in plant tissue and will provide a screen to compare relative termination efficiency between multiple sequences. For these experiments, potentially bicistronic elements are generated containing two distinguishable fluorescent proteins (FP1 and FP2) separated by the transcription terminator to be assessed and IRES. If a particular transcription terminator sequences to be assessed provides efficient transcriptional termination, the expression of FP1 is much greater than the expression of FP2. If a sequence does not terminate efficiently the ratio of FP1:FP2 expression is lower.

**6.1 Vector construction**

Test constructs comprises the following elements, described in order 5' to 3'. A strong constitutive promoter for the target tissue is used to drive expression of the RNA, such as Ubiquitin or ScBV for expression in maize leaf tissue. The most proximal open reading frame encodes for FP1 (*e.g.* DsRed21), followed by restriction sites for insertion of potential transcription terminator sequence elements. Immediately downstream of the insertion site for the transcription terminator sequences to be assessed is the encephalomyocarditis virus (EMCV) IRES element, followed by the open reading frame for FP2 (*e.g.* GFP), and followed by a known plant transcriptional terminator. This downstream terminator needs to be present in order to stabilize transcripts that are not efficiently terminated by the transcription terminator sequence, thereby allowing detection of mRNAs that encode FP2. If *Agrobacterium* mediated transformation experiments will be performed, then the vectors will have to include the LB and RB T-DNA elements flanking the expression cassette and selectable marker genes. (Fig. 15)

**Example 7. System based on inverted repeat of nos terminator**

**7.1 Generation of the positive and negative binary vector control constructs for the screening of terminator activitiy**

To test the transcription termination efficiency of a sequence a construct was generated with a strong constitutive promoter (STPT promoter) upstream of the *nptII* marker gene followed by a short MCS in which putative terminator sequences are cloned and an inverted repeat of the 3´-UTR region from the nos-gene of *Agrobacterium tumefaciens*, with the first repeat element being in the antisense orientation relative to the STPT promoter. This arrangement without a putative terminator sequence serves as a negative control: Both orientations of the nosT are incorporated into the resulting transcript, since the inverted nos element does not lead to transcript truncation and polyadenylation, which gives rise to a transcript with 3´hairpin structure leading to dsRNA-mediated, sequence-specific RNA-degradation and thus silencing of the *nptII*-gene. As a positive control the Ribulose-bisphosphat Carboxlase E9 terminator region is cloned in between the *nptII*-gene and the nos inverted repeat (nos-IR), which leads to the proper termination of *nptII*-transcripts and thus to the expression of *nptII*.

### 7.2 Cloning of pENTR-A1-inv-35s-GFP-E9

Insertion of terminator E9: Lo394-pENTR-A1-inv is opened with *Bam*HI and *Kpn*I. The E9 terminator is isolated from Lo424-pENTR-A1-inv-P2-E9 with *Bam*HI and *Kpn*I and ligated into the opened Lo394 in direct orientation. The resulting construct is named Lo483-pENTR-A1-inv-E9. The orientation of sites is attL4-E9-attR1.

### 7.3 Insertion of the CaMV 35s promoter together with the gene for mGFP5er

The 35s promoter of CaMV and the *Aequorea victoria* gene for the green fluorescent protein mGFP5er-gene were isolated from Lo409-pENTR-35s-GFP5er-GUS with *Hin*dIII/*Sma*I and ligated into Lo483-pENTR-A1-Inv-E9 opened with *Hin*dIII/*Eco*RV in direct orientation. The orientation of sites is attL4-35s-mGFP5er-E9-attR1. The resulting construct is named Lo484-pENTR-A1-inv-35s-GFP-E9 (SEQ ID NO: 9).

### 7.4 Cloning of pENTR-C1-STPT-*nptII*-nosIR

Insertion of the STPT-promoter and the *nptII*-gene: Lo393-pENTR-C1 was opened with *Bam*HI. The STPT promoter with the *nptII*-gene was cut out of Lo441-pENTR1A-STPT-*nptII*-CatpA with *Bam*HI and ligated undirected in the opened Lo393. Clones with the correct orientation were identified via colony-PCR followed by control digests. The orientation of sites is attR2-STPT-*nptII*-attL3. The resulting construct is named Lo485-pENTR-C1-STPT-*nptII*.

### 7.5 Insertion of the terminator nosT in sense orientation

The 257 bp 3´-UTR region of the nopaline synthase (nos) gene from *Agrobacterium tumefaciens* was amplified using PCR from Lo114-pSUN3-Gus-nos with the overhang primers Loy482-NosT-upper-*Sal*I and Loy483-NosT-Lower-*Hin*dIII (SEQ ID NO: 55 and 56).

Loy482-NosT-upper-*Sal*I (SEQ ID NO: 55):
5´-AAATTTGTCGACCGATCGGTCAAACATT-3´

Loy483-NosT-Lower-*Hin*dIII (SEQ ID NO: 56):
5´-AAATTTAAGCTTCCCGATCTAGTAACATAGATGACA-3´

The resulting 282 bp PCR fragment was digested with *Sal*I/*Hin*dIII and cloned in direct orientation into Lo485-pENTR-C1-STPT-*nptII* opened with *Sal*I/*Hin*dIII. The orientation of sites is attR2-STPT-*nptII*-nosT-attL3. The resulting construct is named Lo486-pENTR-C1-STPT-*nptII*-nos.

### 7.6 Insertion of a second nos-terminator element in antisense orientation between the *nptII* gene and the sense nosT together with a stuffer sequence derived from the 3´-region of the gus-gene

The NosT was amplified together with a part of the 3´-region of the gus-gene out of Lo400-pENTR1A (B)-Ln- Prom2-TypDra-nosT. Therefore primers were designed which added a *Sal*I site at the gus sequence and a *Spe*I-site together with a *Bgl*II-site at the end of nosT. Two kinds of constructs were prepared one including a 129 bp gus-spacer sequence between the inverted repeat of nosT and the other consisting of 155 bp spacer between the IR. The shorter version was amplified with the primers Loy494-Gus-upper-*Sal*I-Spacer and Loy492-NosT-lower- *Bgl*II-*Spe*I (SEQ ID NO: 57 and 58).

Loy494- Gus-upper-*Sal*I-Spacer (SEQ ID NO: 57):
5´-TTTTAGTCGACACGCTGGACTGGCATGAACT-3´

Loy492-NosT-lower- *Bgl*II-*Spe*I (SEQ ID NO: 58):
5´-TTTTAAGATCTACTAGTCCGATCTAGTAACATAGATGACA-3´

The longer version was amplified with Loy493-Gus-upper-*Sal*I-Spacer together with Loy492-NosT-lower- *Bgl*II-*Spe*I.

Loy493_Gus_upper_*Sal*I_Spacer (SEQ ID NO: 59):
5´- TTTAAGTCGACAAGTCGGCGGCTTTTCTGCT-3´

Loy492-NosT-lower- *Bgl*II-*Spe*I (SEQ ID NO: 60):
5´-TTTTAAGATCTACTAGTCCGATCTAGTAACATAGATGACA-3´.

The resulting PCR-fragments were digested with *Sal*I/*Bgl*II and ligated into Lo486-pENTR-C1-STPT-*nptII*-nos opened with *Sal*I/*Bgl*II. This resulted in the *nptII* open-reading frame being followed by a nos 3´-UTR in the antisense orientation relative to the STPT promoter, a 129 bp-spacer region respectively 155 bp-spacer region of gus-sequence in the antisense orientation, and a second nosT in the sense orientation. The orientation of sites is attR2-STPT-*nptII*-as nosT-spacer-s nosT-attL3. The resulting constructs were named Lo503a-pENTR-C1-STPT-*nptI*I-IRnos (SEQ ID NO: 8) with the shorter spacer between the IR) and Lo503b-pENTR-C1-STPT-*nptII*-IRnos (SEQ ID NO: 7) with the longer spacer between the IR).

**7.7 Generation of the negative control construct**

**7.7.1 Triple-LR reaction to create the binary expression vector which will serve as the negative control**

The triple-LR-reaction is carried out with the plasmids Lo484-pENTR-A1-inv-35s-GFP-E9, Lo376-pENTR-B2 (without insert; SEQ ID NO: 76) and Lo503a-pENTR-C1-STPT-*nptII*-IRnos, or with Lo503b-pENTR-C1-STPT-*nptII*-IRnos and Lo442-pSUN1-R4R3 (SEQ ID NO: 77), respectively according to the instructions of the manufacturer, using LR plus recombinase mix. The resulting binary plant transformation vectors were named Lo523a-pSUN1-R4-Lo484::Lo376::Lo503a (SEQ ID NO: 6) and Lo523b-pSUN1-R4-Lo484::Lo376::Lo503b (SEQ ID NO: 5), respectively..

**7.8 Generation of the positive control construct**

**7.8.1 Insertion of the E9 terminator upstream of IRnos**

The E9 terminator was isolated from Lo444-pGST-δ-*Kpn*I-LUC$^{+}$ with *Bam*HI/*Eco*RV. Lo503 is opened with *Spe*I, the 5´-protruding ends were completely filled in with *Pfu* turbo polymerase and cut again with *Bgl*II. The *Bgl*II/*Eco*RV fragment of the E9 terminator is ligated into the opened vectors Lo503a and Lo503b (which were digested first with *Spe*I and blunted, and subsequently with *Bam*HI), respectively, in direct orientation. The orientation of sites is attR2-STPT-*nptII*-E9-IRnos-attL3. The resulting constructs were named Lo522a-pENTR-C1-STPT-*nptII*-IRnos (SEQ ID NO: 4) with the shorter spacer between the IR) and Lo522b-pENTR-C1-STPT-*nptII*-IRnos (SEQ ID NO: 3) with the longer spacer between the IR).

### 7.8.2 Triple-LR reaction to create the binary expression vector which will serve as the positive control

The triple-LR-reaction is carried out with the plasmids Lo484-pENTR-A1-inv-35s-GFP-E9, Lo376-pENTR-B2 (without insert; SEQ ID NO: 76) and Lo522a-pENTR-C1-STPT-*nptII*-E9-IRnos or with Lo522b-pENTR-C1-STPT-*nptII*-E9-IRnos and Lo442-pSUN1-R4R3 (SEQ ID NO: 77), respectively, according to the instructions of the manufacturer, using LR plus recombinase mix. The resulting binary plant transformation vectors were named Lo546a-pSUN1-R4-Lo484::Lo376::Lo522a (SEQ ID NO: 2) and Lo546b-pSUN1-R4-Lo484::Lo376::Lo522b (SEQ ID NO: 1), respectively.

## Example 8. System based on inverted repeat of nos terminator using two expression cassettes in head to head orientation

### 8.1 Generation of the positive and negative binary vector control constructs for the screening of terminator activity

To test the transcription termination efficiency of a putative terminator sequence a construct was generated with a strong seed specific promoter (BnAK promoter) upstream of the *ß-Glucuronidase (GUS)* marker gene followed by a short MCS in which putative terminator sequences are cloned and the 3´-UTR region from the nos-gene of *Agrobacterium tumefaciens*. The second expression cassette in this construct contains the promoter of the nos-gene from *Agrobacterium tumefaciens* followed by the *nptII* marker gene and the 3´ nos UTR. Both expression cassettes are oriented in a head to head manner. In transgenic plants this arrangement without a putative terminator sequence serves as a negative control: As the nos terminator is used for both the right hand and left hand expression cassette both orientations of the nosT are incorporated into the resulting transcripts, giving rise to GUS and *nptII* transcripts carrying complentary 3´sequences, leading to hybridization of the two mRNA species, thus resulting in sequence-specific RNA-degradation and total or partial silencing of the ß-Glucuronidase and the *nptII*-gene. As a positive control the Ribulose-bisphosphate Carboxlase E9 terminator region is cloned in between the *GUS*-gene and the nosT, leading to correct truncation of the transcript and enabling high expression of *GUS*. Transcripts from the *nptII* marker gene, carrying the 3´ Tnos UTR are not interacting with the GUS transcripts as there is no complementary sequence present and the *nptII* expression is not impaired (Fig.1 A3).

### 8.2 Cloning of Lo239-pSUN3-GWs-B1-BnAK700::GUS::nosT-B2

Insertion of promoter: The seed specific promoter BnAK700 is isolated by *Hin*dIII/*Bam*HI digestion of Lo229 and inserted into the vector Lo215 pENTR-MCS::GUS::nosT to create Lo235 pENTR-B-BnAK700::GUS::nosT.

Lo 239 (SEQ ID NO: 78) is created by LR reaction of Lo235 with the Gateway destination vector Lo125 pSUN3-GWs-NPTII (Fig. 10).

### 8.3 Cloning of Lo657- pSUN3-GWs-B1-BnAK700::GUS::E9::nosT::B2

Insertion of the E9 Terminator: The Gateway Entry vector Lo235 is cut with *Ecl*136II to create blunt ends. The E9 insert is isolated by restriction of Lo489 with *Eco*RV and *Kpn*I followed by fill in with Klenow fragment. The blunt ended insert is ligated to the linearized vector to create Lo654 pENTR-BnAK700::GUS.E9::nosT.

Lo657 (SEQ ID NO: 79) is created by LR reaction of Lo654 with the Gateway destination vector Lo125 pSUN3-GWs-NPTII (Fig. 10).

**EXAMPLE 9: Development of an *in vivo* screening system to identify terminators by embedding sequences of interest within an intron of a lethal gene or reporter gene.**

A terminator of interest (TOI) is embedded within an intron of a lethal gene including, but not limited to, diphtheria toxin fragment A (DT-A) or a reporter gene including, but not limited to, green fluorescence protein (GFP) (see Fig. 16A). If efficiency of transcription termination is low ("leaky" TOI), there is some transcription of the full length lethal or reporter gene. The intron with the embedded TOI is removed from the transcribed RNA allowing for translation of the full length lethal or reporter protein. In the example using the lethal gene, expression of full length DT-A kills the cells. If the TOI does not allow read through RNA products because of efficient transcription termination ("tight" TOI), only a partial protein is translated and the plant cells are viable. In the example using GFP, a leaky TOI yields full length GFP and cells that fluoresce green. A tight TOI produces only a partial GFP protein and cells don't fluoresce. Control constructs are constructed without a TOI embedded in the intron (Fig. 16B) and with a known terminator, NOS, embedded in the intron (Fig. 16C). See also agenda to Fig. 16 above.

Preferably, a strong constitutive promoter is used for these constructs such as the maize ubiquitin promoter (Zmubi). An octopine synthase terminator (OCS) is added to the end of the cassette to stabilize the read through products. The intron sequence to be used is a potato intron (PIV2) modified here to improve intron splicing efficiency. The modified PIV2 intron contains the following elements to promote efficient intron recognition and splicing in plants (Fig. 17):

(1) Transition at the 5' splice site from moderate AU content (exon) to high AU content (intron).

(2) Transition at the 3' splice site from high AU content (intron) to moderate AU content (exon).

(3) A consensus 5' splice recognition sequence CAG/GUAAGU. '/' identifies the splice site.

(4) A consensus 3' recognition sequence GCAG/G.

(5) A consensus branchpoint sequence CURAY upstream of the 3' splice site.

(6) A polyU tract just downstream of the branchpoint sequence and upstream of the 3' splice site.

A *Bam*HI restriction site is added near the center of the intron for insertion of the transcription terminator of interest (TOI). A *Bam*HI site is compatible with *Sau*3AI and is therefore ideal for insertions of genomic DNA fragments generated by a partial *Sau*3AI digest. The *Bam*HI site can be substituted with other restriction sites to accommodate TOI libraries generated by other means. The cassettes in Fig. 15 can be placed in a binary vector for *Agrobacterium*-mediated plant transformation or in a pUC based vector for biolistic transformation. See also agenda to Fig. 16 above.

**9.1 Vector Construction**

Construct 2 (DT-A version) in Fig. 16 is constructed using the parental vector pTOI03 described above. pTOI03 is digested with *Kpn*I and *Spe*I to remove the GUSint gene but leaving the ZmUbi promoter and OCS terminator. The 3' end of the first half of the DT-A gene is fused to the 5' end of the first half of the PIV2 intron by overlap PCR using the following primers:

JMTOIprim1 (SEQ ID NO: 61)
5'-GGTTCCAAGGTACCAAAACAATGGGCGCTGATGATGTTGTTGAT-3'

JMTOIprim2 (SEQ ID NO: 62)
5'-AAGGTAGAAGCAGAAACTTACCTGGATACGTCACTTTGACCA-3'

JMTOIprim3 (SEQ ID NO: 63)
5'-TGGTCAAAGTGACGTATCCAGGTAAGTTTCTGCTTCTACCTT-3'

JMTOIprim4 (SEQ ID NO: 64)
5'-GGTTCCAAGGATCCATTTATTTTGAAAAAAATATTTG-3'

This overlap PCR places a *Kpn*I site on the 5' end of the DT-A portion of the fused sequences followed by an ATG start codon preceded by the bases AAAACA to enhance translation. It also generates a consensus 5'splice site between the DT-A and intron sequences and a *Bam*HI site 133 bp downstream of the 5' splice site. The 3' end of the second half of the PIV2 intron is fused to the 5' end of the second half of the DT-A gene by overlap PCR using the following primers:

JMTOIprim5 (SEQ ID NO: 65)
5'- GGTTCCAAGGATCCAGTATATAGCAATTGCTTTTC-3'

JMTOIprim6 (SEQ ID NO: 66)
5'- CGAGAACCTTCGTCAGTCCTGCACATCAACAAATTTTGGTCAT
AAAAAAAAAAATATTAGAAAAGTTATAAATTAAAATATAC-3'

JMTOIprim7 (SEQ ID NO: 67)
5'- CTAATATTTTTTTTTTTATGACCAAAATTTGTTGATGTGCAGGA
CTGACGAAGGTTCTCGCAC-3'

JMTOIprim8 (SEQ ID NO: 68)
5'- TTGGAACCACTAGTTTATCGCCTGACACGATTTCCTGC-3'

This overlap PCR places a *Bam*HI site on the 5' end of the PCR product. A tract of 11 consecutive Us at positions +36 to + 26 relative to the 3' splice site and 2 bases downstream of a natural CTAAT consensus branchpoint sequence in the PIV2 intron is added as well as a consensus 3' splice site between the PIV2 and DT-A sequences. This overlap PCR also generates a TAA stop codon at the end of the DT-A open reading frame followed by a *Spe*I restriction enzyme site. The first overlap PCR product is digested with *Kpn*I and *Bam*HI, the second PCR product is digested with *Bam*HI and SpeI, and both PCR products are ligated to pTOI03 digested with *Kpn*I and *Spe*I in a 3-way ligation to make pJMTOI1 (SEQ ID NO: 71). The construct comprises the following features:

| Feature | Location (base) |
|---|---|
| ZmUbi promoter | (1)..(1988) |
| DT-A 5' end of coding sequence | (2007)..(2268) |
| Intron | (2269)..(2488) |
| DTA-A 3' end of coding sequence | (2489)..(2811) |
| OCS terminator | (2818)..(3030) |

Construct 3 (DT-A version) in Fig. 16C will be constructed by first placing *Bam*HI sites on the ends of the NOS terminator using the following PCR primers:

JMTOIprim9 (SEQ ID NO: 69)
5'- GGTTCCAAGGATCCGATCGTTCAAACATTTGGCAA-3'

JMTOIprim10 (SEQ ID NO: 70)
5'- GGTTCCAAGGATCCGATCTAGTAACATAGATGACA-3'

This PCR product is digested with *Bam*HI and ligated into the unique *Bam*HI site within the PIV2 intron of pJMTOI1. Plasmids generated from the ligation are screened to identify those with the correct NOS orientation, yielding pJMTOI2 (SEQ ID NO: 72). The construct comprises the following features:

| Feature | Location (base) |
|---|---|
| ZmUbi promoter | (1)..(1988) |
| DT-A 5' end of coding sequence | (2007)..(2268) |
| Intron 5' end | (2269)..(2380) |
| Nos terminator | (2387)..(2639) |
| Intron 3' end | (2646)..(2747) |
| DTA-A 3' end of coding sequence | (2748)..(3070) |
| OCS terminator | (3077)..(3289) |

**EXAMPLE 10: An *in vivo* selection of efficient terminators.**

In this example, a terminator of interest (TOI) is placed between a reporter gene such as the green fluorescence protein (GFP) and a sequence with little or no homology to plant genes and that is a target of dsRNA mediated RNA silencing (Fig. 18). Each of these elements is under control of a single promoter, in this example, the maize ubiquitin promoter (ZmUbi). When expressed in plants, and if the TOI does not terminate transcription (leaky TOI); the entire transcript (including the region encoding GFP) is degraded by RNA silencing. If the TOI is functional as a terminator, the RNA will not be a target of RNA silencing and GFP will be produced leading to plants that fluoresce green (Fig. 18). TOIs may be obtained by fragmentation of genomic DNA or by a more selective procedure.

A *Bam*HI restriction site will be placed at the junction between GFP and the spacer in construct 3 of Fig. 18 for insertion of the TOI (construct 1) or the NOS terminator (construct 2; pJMTOI3, SEQ ID NO: 73). A *Bam*HI site is compatible with *Sau*3AI and is therefore ideal for insertions of genomic DNA fragments generated by a partial *Sau*3AI digest. The *Bam*HI site can be substituted with other restriction sites to accommodate TOI libraries generated by other means. The cassettes in Fig. 18 can be placed in a

binary vector for *Agrobacterium*-mediated plant transformation. Construct 2 (pJMTOI3) comprises the following features:

| Feature | Location (base) |
|---|---|
| ZmUbi promoter | (1)..(1988) |
| GFP | (2001)..(2696) |
| NOS terminator | (2703)..(2955) |
| Spacer | (2962)..(3161) |
| RNAi target | (3162)..(3461) |
| OCS terminator | (3468)..(3680) |

To perform the terminator screen described in this example, a plant line must be established that can effectively silence RNAs that contain the RNAi target region. In this example, *Arabidopsis* will be used as the host plant although the screen can be used in any plant species that can be transformed. A plant that can effectively silence can be obtained with the following steps using established transformation and genetic screening techniques.

1) Wild-type *Arabidopsis* is transformed with construct 3 in Fig. 18 (pJMTOI4, SEQ ID NO: 74). A $T_1$ plant is selected that is single copy for construct 3 and has strong green fluorescence. Construct 3 (pJMTOI4) comprises the following features:

| Feature | Location (base) |
|---|---|
| ZmUbi promoter | (1)..(1988) |
| GFP | (2001)..(2696) |
| Spacer | (2703)..(2902) |
| RNAi target | (2903)..(3202) |
| OCS terminator | (3209)..(3421) |

2) This fluorescent $T_1$ plant is transformed with construct 4 in Fig. 18 (pJMTOI5; SEQ ID NO: 75). A $T_1$ plant from this transformation is selected that is single copy and hemizygous for construct 3 and single copy for construct 4, and that no longer fluoresces green (silencing plant). Construct 4 (pJMTOI5) comprises the following features:

| Feature | Location (base) |
|---|---|
| ZmUbi promoter | (1)..(1988) |
| RNAi target sense | (2001)..(2300) |
| Spacer | (2309)..(2595) |
| RNAi target anti-sense | (2602)..(2901) |
| NOS terminator | (2908)..(3160) |

3) A $T_2$ plant with respect to construct 4 (pJMTOI5) is obtained that is homozygous for construct 4 and null for construct 3. To generate additional silencing plants, $T_3$ plants can be obtained from the plant isolated in step 3.

After an *Arabidopsis* silencing line containing construct 4 has been established, this line can be transformed with plasmid TOI libraries containing construct 1 and with the control constructs 2 and 3 (pJMTOI3 and (pJMTOI4, respectively) in Fig. 18. $T_1$ plants that fluoresce to a similar extent, as plants transformed with construct 2 will be selected for further analysis. If a selected plant has a single integrant, quantitative RT-PCR targeting GFP and the RNAi target region can determine if the experimental TOI is acting as an efficient terminator. RT-PCR of plants transformed with construct 2 (pJMTOI3)

and construct 3 (pJMTOI4) would serve as controls. If a selected plant has multiple integrants of construct 1, single integrants can be obtained from the $T_2$ generation. The TOIs from selected plants can be amplified and cloned by PCR.

**Example 11: Assays for identifying terminators of interest**

These experiments are performed by bombardment of plant tissues or culture cells (Example 11.1), by PEG-mediated (or similar methodology) introduction of DNA to plant protoplasts (Example 11.2), or by *Agrobacterium*-mediated transformation (Example 11.3). The target tissue for these experiments can be plant tissues (*e.g.* leaf tissue has been described to best support IRES-mediated translation (Urwin *et al.*, 2000), cultured plant cells (*e.g.* maize BMS), or plant embryos for *Agrobacterium* protocols.

The sequences used as potential transcription terminator sequences can either be derived from the *in silico* transcription terminator sequence screen or from a library of random genomic fragments. Ratio of expression of two different reporter genes is measured by quantification of expression of reporter genes or RT-PCR using the protocols in the art in order to determine potentially strong and tight terminator candidates.

**11.1 Transient assay using microprojectile bombardment**

The plasmid constructs are isolated using Qiagen plasmid kit (cat# 12143). DNA is precipitated onto 0.6 µM gold particles (Bio-Rad cat# 165-2262) according to the protocol described by Sanford *et al.* (1993) and accelerated onto target tissues (*e.g.* two week old maize leaves, BMS cultured cells, *etc.*) using a PDS-1000/He system device (Bio-Rad). All DNA precipitation and bombardment steps are performed under sterile conditions at room temperature.

Two mg of gold particles (2 mg/3 shots) are resuspended in 100% ethanol followed by centrifugation in a Beckman Microfuge 18 Centrifuge at 2000 rpm in an Eppendorf tube. The pellet is rinsed once in sterile distilled water, centrifuged, and resuspended in 25 µL of 1 µg/µL total DNA. The following reagents are added to the tube: 220 µL $H_20$, 250 µL 2.5M $CaCl_2$, 50µL 0.1M spermidine, free base. The DNA solution is briefly vortexed and placed on ice for 5 min followed by centrifugation at 500 rpm for 5 min in a Beckman Microfuge 18 Centrifuge. The supernatant is removed. The pellet is resuspended in 600 µL ethanol followed by centrifugation for 1 min at 14,000 rpm. The final pellet is resuspended in 36 µL of ethanol and used immediately or stored on ice for up to 4 hr prior to bombardment. For bombardment, two-week-old maize leaves are cut in approximately 1 cm in length and located on 2 inches diamenter sterilized Whatman filter paper. In the case of BMS cultured cells, 5 mL of one-week-old suspension cells are slowly vacuum filtered onto the 2 inches diameter filter paper placed on a filter unit to remove excess liquid. The filter papers holding the plant materials are placed on osmotic induction media (N6 1-100-25, 0.2 M mannitol, 0.2 M sorbitol) at 27°C in darkness for 2-3 hours prior to bombardment. A few minutes prior to shooting, filters are removed from the medium and placed onto sterile opened Petri dishes to allow the calli surface to partially dry. To keep the position of plant materials, a sterilized wire mesh screen is laid on top of the sample. Each plate is shot with 10 µL of gold-DNA solution once at 2,200 psi for the leaf materials and twice at 1,100 psi for the BMS cultured cells. Following bombardment, the filters holding the samples are transferred onto MS

basal media and incubated for 2 days in darkness at 27°C prior to transient assays. Transient expression levels of the reporter gene are determined quantification of expression of reporter genes or RT-PCR using the protocols in the art in order to determine potentially strong and tight terminator candidates.

**11.2 Transient assay using protoplasts**

Isolation of protoplasts is conducted by following the protocol developed by Sheen (1990). Maize seedlings are kept in the dark at 25°C for 10 days and illuminated for 20 hours before protoplast preparation. The middle part of the leaves are cut to 0.5 mm strips (about 6 cm in length) and incubated in an enzyme solution containing 1% (w/v) cellulose RS, 0.1% (w/v) macerozyme R10 (both from Yakult Honsha, Nishinomiya, Japan), 0.6 M mannitol, 10 mM Mes (pH 5.7), 1 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM β-mercaptoethanol, and 0.1% BSA (w/v) for 3 hr at 23°C followed by gentle shaking at 80 rpm for 10 min to release protoplasts. Protoplasts are collected by centrifugation at 100 x g for 2 min, washed once in cold 0.6 M mannitol solution, centrifuged, and resuspended in cold 0.6 M mannitol ($2 \times 10^6$/mL).

A total of 50 µg plasmid DNA in a total volume of 100 µL sterile water is added into 0.5 mL of a suspension of maize protoplasts ($1 \times 10^6$ cells/mL) and mix gently. 0.5 mL PEG solution (40 % PEG 4000, 100 mM $CaNO_3$, 0.5 mannitol) is added and pre-warmed at 70°C with gentle shaking followed by addition of 4.5 mL MM solution (0.6 M mannitol, 15 mM $MgCl_2$, and 0.1 % MES). This mixture is incubated for 15 minutes at room temperature. The protoplasts are washed twice by pelleting at 600 rpm for 5 min and resuspending in 1.0 mL of MMB solution [0.6 M mannitol, 4 mM Mes (pH 5.7), and brome mosaic virus (BMV) salts (optional)] and incubated in the dark at 25°C for 48 hr. After the final wash step, collect the protoplasts in 3 mL MMB medium, and incubate in the dark at 25 °C for 48 hr. Transient expression levels of the reporter gene are determined quantification of expression of reporter genes or RT-PCR using the protocols in the art in order to determine potentially strong and tight terminator candidates.

**11.3 *Agrobacterium*-mediated transformation in dicotyledonous and monocotyledonous plants**

**11.3.1 Transformation and regeneration of transgenic *Arabidopsis thaliana* (Columbia) plants**

To generate transgenic *Arabidopsis* plants, *Agrobacterium tumefaciens* (strain C58C1 pGV2260) is transformed with the various vector constructs described above. The agrobacterial strains are subsequently used to generate transgenic plants. To this end, a single transformed *Agrobacterium* colony is incubated overnight at 28°C in a 4 mL culture (medium: YEB medium with 50 µg/mL kanamycin and 25 µg/mL rifampicin). This culture is subsequently used to inoculate a 400 mL culture in the same medium, and this is incubated overnight (28°C, 220 rpm) and spun down (GSA rotor, 8,000 rpm, 20 min). The pellet is resuspended in infiltration medium (1/2 MS medium; 0.5 g/L MES, pH 5.8; 50 g/L sucrose). The suspension is introduced into a plant box (Duchefa), and 100 ml of SILWET L-77 (heptamethyltrisiloxan modified with polyalkylene oxide; Osi Specialties Inc., Cat. P030196) is added to a final concentration of 0.02%. In a desiccator, the plant box with 8 to 12 plants is exposed to a vacuum for 10 to 15 minutes, followed by spontaneous aeration. This is repeated twice or 3 times. Thereupon, all plants are planted into flowerpots with moist soil and grown under long-

day conditions (daytime temperature 22 to 24°C, nighttime temperature 19°C; relative atmospheric humidity 65%). The seeds are harvested after 6 weeks.

As an alternative, transgenic *Arabidopsis* plants can be obtained by root transformation. White root shoots of plants with a maximum age of 8 weeks are used. To this end, plants which are kept under sterile conditions in 1 MS medium (1% sucrose; 100mg/L inositol; 1.0 mg/L thiamine; 0.5 mg/L pyridoxine; 0.5 mg/L nicotinic acid; 0.5 g MES, pH 5.7; 0.8 % agar) are used. Roots are grown on callus-inducing medium for 3 days (1x Gamborg's B5 medium; 2% glucose; 0.5 g/L mercaptoethanol; 0.8% agar; 0.5 mg/L 2,4-D (2,4-dichlorophenoxyacetic acid); 0.05 mg/L kinetin). Root sections 0.5 cm in length are transferred into 10 to 20 mL of liquid callus-inducing medium (composition as described above, but without agar supplementation), inoculated with 1 mL of the above-described overnight *Agrobacterium* culture (grown at 28°C, 200 rpm in LB) and shaken for 2 minutes. After excess medium has been allowed to run off, the root explants are transferred to callus-inducing medium with agar, subsequently to callus-inducing liquid medium without agar (with 500 mg/L betabactyl, SmithKline Beecham Pharma GmbH, Munich), incubated with shaking and finally transferred to shoot-inducing medium (5 mg/L 2-isopentenyladenine phosphate; 0.15 mg/L indole-3-acetic acid; 50 mg/L kanamycin; 500 mg/L betabactyl). After 5 weeks, and after 1 or 2 medium changes, the small green shoots are transferred to germination medium (1 MS medium; 1% sucrose; 100 mg/L inositol; 1.0 mg/L thiamine; 0.5 mg/L pyridoxine; 0.5 mg/L nicotinic acid; 0.5 g MES, pH 5.7; 0.8% agar) and regenerated into plants.

**11.3.2 Transformation and regeneration of crop plants**

The *Agrobacterium*-mediated plant transformation using standard transformation and regeneration techniques may also be carried out for the purposes of transforming crop plants (Gelvin 1995; Glick 1993). For example, oilseed rape can be transformed by cotyledon or hypocotyl transformation (Moloney 1989; De Block 1989). The use of antibiotics for the selection of *Agrobacteria* and plants depends on the binary vector and the *Agrobacterium* strain used for the transformation. The selection of oilseed rape is generally carried out using kanamycin as selectable plant marker. The *Agrobacterium*-mediated gene transfer in linseed (*Linum usitatissimum*) can be carried out using for example a technique described by Mlynarova (1994). The transformation of soybean can be carried out using, for example, a technique described in EP–A1 0 424 047 or in EP–A1 0 397 687, US 5,376,543, US 5,169,770. The transformation of maize or other monocotyledonous plants can be carried out using, for example, a technique described in US 5,591,616. The transformation of plants using particle bombardment, polyethylene glycol-mediated DNA uptake or via the silicon carbonate fiber technique is described, for example, by Freeling & Walbot (1993) "The maize handbook" ISBN 3-540-97826-7, Springer Verlag New York).

## REFERENCES

The references listed below and all references cited herein are incorporated herein by reference to the extent that they supplement, explain, provide a background for, or teach methodology, techniques, and/or compositions employed herein.

1. Abremski *et al.* (1986) J. Biol. Chem. 261(1):391
2. Allen *et al.* (1996) Plant Cell 8(5), 899-913
3. Allen *et al.* (2000) Plant Mol Biol 43(2-3):361-76
4. An *et al.* (1985) EMBO J. 4:277-287
5. Anandalakshmi *et al.* (1998) Proc Natl Acad Sci USA 95(22):13079-84
6. Andersen *et al.* (1989) Arch Microbiol 152:115-118
7. Anderson & Young (1985) Quantitative Filter Hybridization, in Nucleic Acid Hybridization
8. Angell *et al.* (1999) Plant J 20(3):357-362
9. Araki *et al.* (1992) J. Mol. Biol. 225(1):25
10. Argos *et al.* (1986) EMBO J. 5:433-440
11. Atanassova *et al.* (1992) Plant J 2(3): 291-300
12. Ausubel *et al.* (1995) Short Protocols in Molecular Biology, 3.sup.rd Edition (John Wiley & Sons
13. Bailey *et al.*, "Manipulation of Baculovirus Vectors," in Methods in Molecular Biology, Volume 7: Gene Transfer and Expression Protocols, Murray (ed.), p.147-168 (The Humana Press, Inc. 1991)
14. Bartel & Szostak (1993) Science 261:1411-1418
15. Bäumlein *et al.* (1991a) Mol Gen Genet 225(3):459-467
16. Bäumlein *et al.* (1991b) Mol Gen Genet 225:121-128
17. Bayley *et al.* (1992) Plant Mol Biol. 18(2):353-361
18. Beclin *et al.* (1993) Transgenics Res 2:4855
19. Beerli *et al.* (2000) Proc Natl Acad Sci USA 97(4):1495-500
20. Beerli *et al.* (1998) Proc Natl Acad Sci USA 95(25):14628- 14633
21. Beerli *et al.* (2000) J Biol Chem 275(42):32617-32627
22. Belfort & Roberts (1997) Nucleic Acids Res 25:3379-3388
23. Bennett (1990) in Goodman and Gilman: the Pharmacological Basis of Therapeutics. 8th ed. eds. Gilman AG et al. (Pergamon Press, New York) pp. 1165-1181
24. Benoist *et al.* (1981) Nature 290:304
25. Besnard *et al.* (1987) Mol Cell Biol 7:4139
26. Bevan *et al.* (1984) Nucl Acid Res 12,8711-8720
27. Bidney *et al.* (1992) Plant Mol. Biol. 18:301-313
28. Blanc *et al.* (1996) Biochimie 78(6):511-517
29. Bonning *et al.* (1994) J. Gen. Virol. 75:1551 (1994)
30. Broach *et al.* (1982) Cell 29:227-234
31. Brown (1991) (ed.), Molecular Biology Labfax Academic Press
32. Bruce *et al.* (1989) Proc Natl Acad Sci USA 86:9692-9696
33. Brummell *et al.* (2003) Plant J. 33:793-800
34. Bustos *et al.* (1989) Plant Cell 1(9):839-53
35. Calabrisi & Chabner (1990) in Goodman and Gilman: the Pharmacological Basis of Therapeutics. 8th ed., eds. Gilman et al. (Pergamon Press, New York) pp. 1209-1263
36. Campbell (1992) J. Bacteriol. 174(23):7495
37.. Canaday *et al.* (1992) Mol Gen Genet 235:292-303
38. Cecchini *et al.* (1998) Mutat Res 401(1-2):199-206
39. Chasin *et al.* (1986) Som. Cell. Mol. Genet. 12:555
40. Chazenbalk & Rapoport (1995) J. Biol. Chem. 270:1543
41. Chen & Winans (1991) J. Bacteriol. 173: 1139-1144

42. Christensen *et al.* (1989) Plant Mol. Biol. 12: 619-632
43. Christensen *et al.* (1992) Plant Mol Biol 18:675-689
44. Christou *et al.* (1988) Plant Physiol 87:671-674
45. Christou *et al.* (1991) Bio/Technology 9:957-962
46. Chui *et al.* (1996) Curr Biol 6:325-330
47. Coombs (1994) Dictionary of Biotechnology, Stockton Press, New York N.Y.]
48. Corneille *et al.* (2001) Plant J 27:171-178
49. Cornell *et al.* (1996) 317:285-290).
50. Cramer *et al.* (1999) Current Topics in Microbiology and Immunology 240:95-118
51. Cramer *et al.* (2001) FEBS Letters 498:179-182
52. Czako & Marton (1994) Plant Physiol 104:1067-1071
53. Dale & Ow (1991) Proc Nat'l Acad Sci USA 88:10558-10562
54. Damon *et al.* (1989) Pharmac Ther 43:155-189)
55. Datta *et al.* (1990b) Bio/Technology 8:736-740
56. DE 10212892
57. de Block *et al.* (1987) EMBO J 6:2513-2518
58. De Block *et al.* (1989) Plant Physiol. 91:694-701)
59. de Feyter *et al.* (1996) Mol Gen Genet. 250(3):329-338
60. De la Pena *et al.* (1987) Nature 325:274-276
61. Deblaere *et al.* (1985) Nucl Acids Res 13:4777-4788
62. Della-Cioppa *et al.* (1987) Bio/Technology 5:579-584
63. Dellaporta *et al.* (1988) In: Chromosome Structure and Function: Impact of New Concepts, 18th Stadler Genetics Symposium, 11:263-282
64. Depicker *et al.* (1988) Plant Cell rep 104:1067-1071
65. Dietert *et al.* (1982) Plant Science Letters 26:233-240
66. Donald *et al.* (1996) J Biol Chem 271(24):14010-14019
67. Dotson *et al.* (1996) J Biol Chem 271(42): 25754-25761
68. Dotson *et al.* (1996) Plant J 10(2):383-392
69. Dreier *et al.* (2000) J Mol Biol 303(4):489-502
70. Dreier *et al.* (2001) J Biol Chem 276(31):29466-78
71. Drocourt *et al.* (1990) Nucl. Acids Res. 18:4009
72. Du *et al.* (1989) Genet Manip Plants 5:8-12
73. Dunahay *et al.* (1995) J Phycol 31:10004-1012
74. Dunwell (2000) J Exp Bot 51 Spec No:487-96
75. Ebinuma *et al.* (2000a) Proc Natl Acad Sci USA 94:2117-2121
76. Ebinuma *et al.* (2000b) Selection of Marker-free transgenic plants using the oncogenes (*ipt, rol* A, B, C) of *Agrobacterium* as selectable markers, In Molecular Biology of Woody Plants. Kluwer Academic Publishers
77. Eichholtz *et al.* (1987) Somatic Cell and Molecular Genetics 13, 67-76
78. Ellegren & Låås (1989) J. Chromatogr. 467, 217
79. EP-A1 0 120 516
80. EP-A1 0 270 615
81. EP-A1 0 321 201
82. EP-A1 0 333 033
83. EP-A1 0 335 528
84. EP-A1 0 388 186
85. EP-A1 0 397 687
86. EP-A1 0 424 047
87. EP-A1 0 595 837
88. EP-A1 0 595 873
89. EP-A1 0 601 092

90. EP-A1 0 807 836
91. EP-A1 0 291 533
92. EP-A1 0 360 257
93. Erikson *et al.* (2004) Nat Biotechnol. 22(4):455-8
94. Etcheverry (1996) "Expression of Engineered Proteins in Mammalian Cell Culture," in Protein Engineering: Principles and Practice, Cleland *et al.* (eds.), p.163-181 (John Wiley & Sons, Inc.
95. Fagard & Vaucheret (2000) Plant Mol Biol 43(2-3):285-93
96. Falciatore *et al.* (1999) Marine Biotechnology 1(3):239-251
97. Famulok & Mayer (1999) Curr Top Microbiol Immunol 243:123-36
98. Fedoroff & Smith (1993) Plant J 3:273- 289
99. Fire *et al.* (1998) Nature 391:806-811
100. Foecking *et al.* (1980) Gene 45:101
101. Fenwick (1985) The HGPRT system, in *Molecular Cell Genetics* 1st Ed. (ed Gottesman, M.) Wiley, New York, pp.333-373
102. Fraley *et al.* (1982) Proc. Natl. Acad. Sci. USA 79:1859-1863
103. Fraley *et al.* (1983) Proc Natl Acad Sci USA 80:4803
104. Fraley *et al.* (1985) CRC Crit. Rev. Plant. Sci., 4:1-45
105. Franck *et al.* (1980) Cell 21:285-294
106. Franken *et al.* (1997) Curr Opin Biotechnol 8(4):411-416
107. Freeling & Walbot (1994) The Maize Handbook, Chapter 116, Eds., Springer, New York
108. Fromm *et al.* (1986) Nature 319:791-793
109. Fromm *et al.* (1990) Bio/Technology 8:833-839
110. Gallego (1999) Plant Mol Biol 39(1):83-93
111. Gallie *et al.* (1987) Nucl Acids Res 15:8693-8711
112. Gardner *et al.* (1986) Plant Mol Biol 6:221-228
113. Gatignol *et al.* (1987) Mol. Gen. Genet. 207:342
114. Gatz *et al.* (1991) Mol Gen Genetics 227:229-237
115. Gatz *et al.* (1992) Plant J 2:397-404
116. Gatz *et al.* (1994) Mol Gen Genetics 243:32-38).
117. Gatz *et al.* (1997) Annu Rev Plant Physiol Plant Mol Biol 48:89-108
118. Gaudin & Jouanin (1995) Plant Mol Biol. 28(1):123-36
119. Gautier *et al.* (1987) Nucleic Acids Res 15:6625-6641)
120. Gavilondo & Larrick (2000) Biotechniques 29(1):128-138
121. Gelvin *et al.* (Eds) (1990) Plant Molecular Biology Manual; Kluwer Academic Publisher, Dordrecht, The Netherlands
122. Gelvin & Schilperoort (1995) Plant Molecular Biology Manual, 2nd Edition, Dordrecht: Kluwer Academic Publ., ISBN 0-7923-2731-4
123. GenBank Acc. No.: AB025109
124. GenBank Acc. No.: U38846
125. GenBank Acc. No.: X03677
126. GenBank Acc.-No.: AB016260 (Protein_id="BAA87807.1)
127. GenBank Acc.-No.: AE009419
128. GenBank Acc.-No.: AF212863
129. GenBank Acc.-No.: AC079674
130. GenBank Acc.-No.: J01603
131. GenBank Acc.-No.: M13422
132. GenBank Acc.-No.: M60917
133. GenBank Acc.-No.: M61151
134. GenBank Acc.-No.: AF039169
135. GenBank Acc.-No.: AB025110

136. GenBank Acc.-No.: NC002147
137. GenBank Acc.-No.: U02443
138. GenBank Acc.-No.: U10247
139. GenBank Acc.-No.: U44852
140. GenBank Acc.-No.: X00221
141. GenBank Acc.-No.: X77943
142. GenBank Acc.-No.: M12196
143. GenBank Acc.-No.: AF172282
144. GenBank Acc.-No.: X04049
145. GenBank Acc.-No.: AF253472
146. GenBank Acc.-No: J02224
147. GenBank Acc.-No.: V00470
148. GenBank Acc.-No.: V00467
149. GenBank Acc.-No: M26950.
150. GenBank Acc.-No: M32238.
151. GenBank Acc.-No: NC_003308 (Protein_id="NP_536128.1),
152. GenBank Acc.-No: S56903,
153. GeneBank Acc.-No.: U60066
154. Gleave *et al.* (1999) Plant Mol Biol 40(2):223-235
155. Gleeson *et al.* (1986) J. Gen. Microbiol. 132:3459
156. Glick & Thompson (1993) Methods in Plant Molecular Biology and Biotechnology, Boca Raton: CRC Press, ISBN 0-8493-5164-2)
157. Glimelius (1984) Physiol Plant 61:38-44
158. Gordon-Kamm *et al.* (1990) Plant Cell 2:603-618
159. Goring *et al.* (1991) Proc. Nat'l Acad. Sci. USA 88:1770-1774
160. Gorman *et al.* (1982) Proc. Nat'l Acad. Sci. USA 79:6777
161. Gottesman (1985) Molecular Cell Genetics, John Wiley and Sons, New York
162. Grant *et al.* (1995) Science 269, 843-846
163. Greener & Callahan (1994) Strategies 7:32-34
164. Griesbach (1992) HortScience 27:620
165. Gruber *et al.* (1993) "Vectors for Plant Transformation," in METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY; pp.89-119
166. Grussenmeyer *et al.* (1985) Proc. Nat'l Acad. Sci. 82:7952
167. Guerrero *et al.* (1993) Mol Gen Genet 224:161-168
168. Hajdukiewicz *et al.* (1994) Plant Mol Biol 25:989-994
169. Hamer *et al.* (1982) J. Molec. Appl. Genet. 1:273
170. Hansen *et al.* (1994) Proc. Natl. Acad. Sci. USA 91:7603-7607
171. Hardy (1985) "Bacillus Cloning Methods," in DNA Cloning: A Practical Approach, Glover (ed.) (IRL Press)
172. Hasegawa *et al.* (2003) The Plant journal 33:1063-1072
173. Haselhoff & Gerlach (1988) Nature 334:585-591
174. Haseloff *et al.* (1997) Proc Natl Acad Sci USA 94(6):2122-2127
175. Hashida-Okado *et al.* (1998) FEBS Letters 425:117
176. Hayakawa *et al.* (1992) Proc Natl Acad Sci USA 89:9865-9869
177. Hayford *et al.* (1988) Plant Physiol. 86:1216
178. Hershey *et al.* (1991) Mol Gen Genetics 227:229-237
179. Hille *et al.* (1986) Plant Mol. Biol. 7:171
180. Hill-Perkins & Possee (1990) J. Gen. Virol. 71:971
181. Hoekema (1985) In: The Binary Plant Vector System, Offsetdrukkerij Kanters B.V., Alblasserdam, Chapter V
182. Hoekema *et al.* (1983) Nature 303:179-181

183. Hoess & Abremski (1990) In Nucleic Acids and Molecular Biology, vol. 4. Eds.: Eckstein and Lilley, Berlin-Heidelberg: Springer-Verlag; pp. 90-109
184. Hoess *et al.* (1986) Nucleic Acids Research 14(6):2287
185. Holliger & Bohlen (1999) Cancer & Metastasis Reviews 18(4):411-419
186. Holsters *et al.* (1978) Mol Gen Genet 163:181-187
187. Holtorf *et al.* (1995) Plant Mol Biol 29:637-649
188. Hood & Jilka (1999) Curr Opin Biotechnol. 10(4):382-6
189. Hood *et al.* (1986) J Bacteriol 168:1291-1301
190. http://rebase.neb.com/rebase/ rebase.homing.html;
191. http://www.biomedcentral.com/1471-2180/1/15
192. Inze *et al.* (1984) Mol Gen Genet 194:265-274
193. Jacobs *et al.* (1988) Biochem Genet 26(1-2):105-22
194. Janssen (1989) J Bacteriol 171(12):6791-9)
195. Janssen *et al.* (1994) Annu Rev Microbiol 48: 163-191
196. Jasin (1996) Trends Genet 12:224-228
197. Bennett (1990) Chapter 50: Antifungal Agents, in Goodman and Gilman's the Pharmacological Basis of Therapeutics 8th ed., A.G. Gilman, ed., Pergamon Press, New York
198. Jefferson (1987b) Plant Mol. Bio. Rep. 5:387-405
199. Jefferson *et al.* (1987a) EMBO J. 6:3901-3907
200. Jenes *et al.* (1983) Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, edited by Kung & Wu, Academic Press 128-143
201. Jolly *et al.* (1983) Proc Natl Acad Sci USA 80:477
202. Jones *et al.* (1987) Mol Gen Genet 210:86
203. Jorgensen *et al.* (1996) Plant Mol Biol 31(5):957-973
204. Joseffson *et al.* (1987) J Biol Chem 262:12196-12201
205. Kado (1991) Crit Rev Plant Sci 10:1
206. Kang & Kim (2000) J Biol Chem 275(12):8742-8748
207. Karlin-Neumannn *et al.* (1991) Plant Cell 3:573-582
208. Kasuga *et al.* (1999) Nature Biotechnology 17:276-286
209. Kaufman (1990a) Meth. Enzymol. 185:487
210. Kaufman (1990b) Meth. Enzymol. 185:537
211. Kaufman *et al.* (1991) Nucl. Acids Res. 19:4485
212. Kavanagh (2002) Plant J. 32, 391-400
213. Keown (1990) Methods in Enzymology 185:527-537
214. Kilstrup *et al.* (1989) J Bacteriol 171:2124-2127
215. Kim *et al.* (2003) Biotechnology Progress 19:1620-1622
216. Kim *et al.* (1997) Proc Natl Acad Sci USA 94(8):3616 -3620
217. Klapwijk *et al.* (1980) J. Bacteriol., 141,128-136
218. Klug (1999) J Mol Biol 293(2):215-218
219. Knoll *et al.* (1998) Mol Cell Biol 18(2):807-814
220. Kobayashi *et al.* (1995) Jpn J Genet 70(3):409-422
221. Koechlin *et al.* (1966) Biochemical Pharmacology 15:434-446
222. Koncz & Schell (1986) Mol Gen Genet 204:383-396
223. Koprek *et al.* (1999) Plant J 19(6):719-726
224. Krens *et al.* (1982) Nature 296:72-74
225. Kuersten & Goodwin (2003) Nature Reviews Genetics 4:626-637
226. Landy (1989) Ann. Rev. Biochem. 58:913-949
227. Landy (1993) Curr Opin Genet Dev. 3(5):699-707
228. Last *et al.* (1991) Theor. Appl. Genet. 81:581-588
229. Lazzeri (1984) Annals of Botany 54:341-350
230. Leffel *et al.* (1997) Biotechniques. 23(5):912-8

231. Lepetit *et al.* (1992) Mol. Gen. Genet. 231:276-285
232. Li *et al.* (1992) Plant Mol Biol 20:1037-1048
233. Li (1982) Plant Cell Rep 1:209-211
234. Liu *et al.*, (2002) Analytical Biochemistry 300:40-45
235. Lloyd & Davis *et al.* (1994) Mol Gen Genet. 242(6):653-657
236. Lucknow (1996) "Insect Cell Expression Technology," in Protein Engineering: Principles and Practice, Cleland *et al.* (eds.), pages 183-218 (John Wiley & Sons, Inc.).
237. Luckow *et al.*(1993) J. Virol. 67:4566
238. Ludwig *et al.* (1990) Science 247:449
239. Luo & Wu (1988) Plant Mol. Biol. Rep. 6:165-174
240. Luo *et al.* (1996) Arch. Biochem. Biophys. 329:215
241. Ma & Vine (1999) Curr Top Microbiol. Immunol. 236:275-92
242. Maniatis *et al.* (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor (NY)
243. Mapp *et al.* (2000) Proc Natl Acad Sci USA 97(8):3930-3935
244. Margraff *et al.* (1980) Experimentia 36: 846)
245. Markie (1996) Methods Mol. Biol. 54:359
246. Matzke *et al.* (1994) Mol Gen Genet 244:219-229
247. Matzke *et al.* (2000) Plant Mol Biol 43:401-415
248. Matzke *et al.* (1989) The EMBO Journal 8(3):643-649
249. McElroy *et al.* (1990) Plant Cell 2:163171
250. McKnight (1982) Cell 31:355
251. McKnight *et al.* (1980) Nucl Acids Res 8(24):5931-5948
252. McKnight *et al.* (1980) Nucl Acids Res 8(24):5949-5964
253. Mett *et al.* (1993) PNAS 90: 4567-4571
254. Miki *et al.* (1993) "Procedures for Introducing Foreign DNA into Plants" in METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY; pp.67-88
255. Millar *et al.* (1992) Plant Mol Biol Rep 10:324-414
256. Mlynarova *et al.* (1994) Plant Cell Report 13:282-285
257. Mlynarova *et al.* (2003) Plant Cell. 15(9):2203-17
258. Mlynarova *et al.* (2002) Genetics 160, 727-40
259. Mol *et al.* (1990) FEBS Lett 268(2):427-430
260. Moloney *et al.* (1989) Plant Cell Reports 8:238-242
261. Moreno *et al.* (1997) J. Immunol. 158: 5841-5848
262. Morganti *et al.* (1996) Biotechnol. Appl. Biochem. 23:67
263. Mozo & Hooykaas (1991) Plant Mol. Biol. 16 : 917-918.
264. Mullen *et al.* (1992) Proc Natl Acad Sci USA 89(1):33-37
265. Murai *et al.* (1983) Science 23: 476-482
266. Mzoz & Moolten (1993) Human Gene Therapy 4:589-595
267. Naested *et al.* (1999) Plant J 18(5):571-576
268. Nan *et al.* (1995) In "Biotechnology in Agriculture and Forestry," Ed. Y. P. S. Bajaj, Springer-Verlag Berlin Heidelberg, Vol 34:145-155
269. Napoli *et al.* (1990) The Plant Cell 2:279-289
270. Napoli *et al.* (1990) Plant Cell 2:279-289
271. Nehra *et al.* (1994) Plant J. 5:285-297
272. O'Keefe (1991) Biochemistry 30(2):447-55
273. O'Keefe *et al.* (1994) Plant Physiol 105:473-482
274. Odell *et al.* (1985) Nature 313:810-812
275. Olhoft *et al.* (2001) Plant Cell Rep 20: 706–711
276. Ono *et al.* (1997) Hum Gene Ther 8(17):2043-55
277. Ow *et al.* (1986) Science 234:856-859

278. Owen *et al.* (1992) Biotechnology (N Y) 10(7):790-794
279. Owens *et al.* (1973) Weed Science 21:63-66)
280. Padidam & Cao (2001) BioTechniqus 31:328-334
281. Paszkowski *et al.* (1984) EMBO J 3:2717-2722
282. Patel *et al.* (1995) "The baculovirus expression system," in DNA Cloning 2: Expression Systems, 2nd Edition, Glover et al. (eds.), pages 205-244 (Oxford University Press)
283. Perera *et al.* (1993) Plant Mol. Biol 23(4): 793-799;
284. Perl *et al.* (1996) Nature Biotechnol 14: 624–628
285. Pfeifer *et al.* (1997) Gene 188:183
286. Polak & Scholer (1975) Chemotherapy (Basel) 21:113-130
287. Polak *et al.* (1976) Chemotherapy 22:137-153
288. Potrykus (1991) Ann Rev Plant Physiol Plant Mol Biol 42:205-225
289. Prasher *et al.* (1985) Biochem Biophys Res Commun 126(3):1259-1268
290. Preston *et al.* (1981) J Virol 38(2):593-605
291. Proudfoot (1986) Nature 322:562-565
292. Qian *et al.* (1992) J. Biol. Chem. 267(11):7794
293. Randez-Gil *et al.* (1995) Yeast 11:1233-1240
294. Ratcliff F et al. (2001) Plant J 25(2):237-45
295. Rathore *et al.* (1993) Plant Mol Biol 21(5):871-884
296. Raymond *et al.,* Yeast 14:11-23 (1998)
297. Razik & Quatrano (1997) Plant Cell 9:1791-1803
298. Richardson (ed.) (1995) Baculovirus Expression Protocols (The Humana Press, Inc.)
299. Risseeuw (1997) Plant J 11(4):717-728
300. Roberts & Macelis (2001) Nucl Acids Res 29:268-269
301. Romanos *et al.* (1995) "Expression of Cloned Genes in Yeast," in DNA Cloning 2: Expression Systems, 2.sup.nd Edition, p.123-167 (IRL Press)
302. Rouster *et al.* (1998) Plant J 15:435-440
303. Ruiz (1998) Plant Cell 10(6):937-46)
304. Russel (1999), Current Topics in Microbiology and Immunology 240:119-138
305. Salomon & Puchta (1998) EMBO J 17(20):6086-6095
306. Sambrook *et al.* (1989) Cold Spring Harbor Laboratory Press; ISBN 0-87969-309-6
307. Sanford (1990) Physiologia Plantarium 79:206-209
308. Sanger *et al.* (1977) Proc Natl Acad Sci USA 74:5463-5467
309. Sauer (1994) Current Opinion in Biotechnology 5:521-527
310. Sautter *et al.* (1991) Bio/Technology 9:1080-1085
311. Scheeren-Groot *et al.* (1994) J. Bacteriol 176: 6418-6426
312. Schena *et al.* (1991) Proc Nat'l Acad Sci USA 88:10421
313. Schenborn & Groskreutz (1999) Mol Biotechnol 13(1):29-44
314. Schlaman & Hooykaas (1997) Plant J 11:1377-1385
315. Schroder *et al.* (1984) Eur J Biochem 138:387-391
316. Schwartz (1981) Environ Health Perspect 37:75-7
317. Segal & Barbas 3rd. (2000) Curr Opin Chem Biol 4(1):34-39
318. Sekowska *et al.* (2001) BMC Microbiol 1:15
319. Sengupta-Gopalan *et al.* (1985) Proc. Nat'l Acad. Sci. USA 82: 3320-3324
320. Serino (1997) Plant J 12(3):697-701
321. Shah *et al.* (1986) Science 233: 478
322. Shahla *et al.* (1987) Plant Mole. Biol. 8:291-298
323. Sharrocks *et al.* (1997) Int J Biochem Cell Biol 29(12):1371-1387
324. Sheehy *et al.* (1988) Proc Natl Acad Sci USA 85: 8805-8809
325. Sheen (1990) Plant Cell 2:1027-1038
326. Sheen (1995) Plant J 8(5):777-784

327. Shewmaker *et al.* (1985) Virology 140:281-288
328. Shillito *et al.* (1985) Bio/Technology, 3:1099-1103
329. Shimamoto *et al.* (1989) Nature 338:274-276
330. Shirsat *et al.* (1989) Mol Gen Genet 215(2):326-331
331. Sidorenko *et al.* (2003) Transgenic Research. 12(2):137-54
332. Silhavy *et* al. (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
333. Simpson *et al.* (1985) EMBO J 4:2723-2729
334. Smith *et al.* (1990) Mol Gen Genet 224:447-481
335. Sorscher *et al.* (1994) Gene Therapy 1:233-238
336. Srivastava & Schlessinger (1991) Gene 103:53
337. St. Clair *et al.* (1987) Antimicrob Agents Chemother 31(6):844-849
338. Stalberg *et al.* (1996) Planta 199:515-519
339. Steinecke *et al.* (1992) EMBO J 11(4):1525- 1530
340. Steinecke (1995) Ribozymes, Methods in Cell Biology 50, Galbraith et al. eds, Academic Press, Inc. p.449-460
341. Stief *et al.* (1989) Nature 341:343
342. Stougaard (1993) Plant J 3:755-761
343. Stringham (1979) Z Pflanzenphysiol 92:459-462
344. Sundaresan *et al.* (1995) Gene Develop 9: 1797-1810
345. Sundaresan *et al.* (1995) Gene Develop 9:1797-1810
346. Svab *et al.* (1990) Plant Mol. Biol. 14:197
347. Tanner (1999) FEMS Microbiol Rev 23(3):257-275
348. Taylor *et al.* (1985) "The APRT Systern", pp., 311-332, M. Gottesman (ed.), Molecular Cell Genetics, John Wiley and Sons, New York
349. The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994)
350. Thykjaer *et al.* (1997) Plant Mol Biol 35(4):523-530
351. Tian *et al.* (1997) Plant Cell Rep 16:267-271
352. Timko *et al.* (1985) Nature 318: 579-582
353. Tissier *et al.* (1999) Plant Cell 11:1841-1852
354. Tomashow *et al.* (1984) Proc Natl Acad Sci USA 81:5071-5075
355. Topfer *et al.* (1989) Plant Cell, 1:133-139
356. Tsai *et al.* (1998) Adv Drug Deliv Rev 30(1-3):23-31
357. Tucker & Burke (1997) Gene 199:25
358. Twell *et al.* (1983) Sex. Plant Reprod. 6: 217-224
359. Twell *et al.* (1989b) Mol Gen Genet 217:240-245
*360.* Ulmasov & Folk (1995) The Plant Cell 7:1723-1734
361. Upadhyaya *et al.* (2000) Plant Mol Biol Rep 18:227-223
362. Urwin *et al.* (2000) *Plant J* 24: 583-589
363. US 4,486,533
364. US 4,599,311
365. US 4,615,974
366. US 4,661,454
367. US 4,801,340
368. US 4,845,075
369. US 4,870,008
370. US 4,882,279
371. US 4,931,373
372. US 4,935,349
373. US 4,962,028
374. US 4,975,374

375. US 4,977,092
376. US 4,987,071
377. US 4,990,446,
378. US 4.940.838
379. US 5,037,743
380. US 5,037,743
381. US 5,063,154
382. US 5,139,936
383. US 5,143,830
384. US 5,162,228
385. US 5,169,770
386. US 5,180,873
387. US 5,300,435
388. US 5,334,575
389. US 5,352,605
390. US 5,358,866;
391. US 5,376,543
392. US 5,426,041
393. US 5,501,967
394. US 5,504,200
395. US 5,584,807
396. US 5,591,616
397. US 5,608,152
398. US 5,683,439
399. US 5,716,808
400. US 5,736,383,
401. US 5,888,732
402. US 5,034,323.
403. US 5,116,742
404. US 5,254,801
405. Vagner *et al.* (2001) EMBO Rep. 2(10):893-8
406. Van der Krol *et al.* (1990) Plant Cell 2:291-99
407. Van Laerebeke *et al.* (1974) Nature 252:169-170
408. Vancanneyt *et al.* (1990) Mol Gen Genet 220(2):245-250
409. Vanden Elzen *et al.* (1985) Plant Mol Biol. 5:299
410. VanOnckelen *et al.* (1986) FEBS Lett. 198: 357-360
411. Vasil *et al.* (1992) Bio/Technology, 10:667-674
412. Velten *et al.* (1984) EMBO J. 3(12): 2723-2730
413. Villemure *et al.* (2001) J. Mol. Biol. 312, 963-974
414. Wagner *et al.* (1981) Proc Natl Acad Sci USA 78(3):1441-1445
415. Wan & Lemaux (1994) Plant Physiol. 104:3748
416. Ward *et al.* (1993) Plant. Mol. Biol. 22:361-366
417. Waterhouse *et al.* (1998) Proc Natl Acad Sci USA 95:13959-64
418. Watson *et al.* (1975) J. Bacteriol 123:255-264
419. Watson *et al.* (1985) EMBO J 4(2):277-284
420. Weeks *et al.* (1993) Plant Physiol 102:1077-1084
421. Whitelam (1996) Trend Plant Sci 1:286-272)
422. Wigler *et al.* (1977) Cell 11(1):223-232;
423. Wigler *et al.* (1979) Proc Natl Acad Sci USA 76(3):1373-6
424. Wisman *et al.* (1991) Mol Gen Genet 226(1-2):120-8
425. WO 00/26388

426. WO 00/44895
427. WO 00/44914
428. WO 00/49035
429. WO 00/63364
430. WO 00/68374
431. WO 02/00900
432. WO 03/060133
433. WO 03/078629
**434.** WO 03/052104
435. WO 84/02913
436. WO 91/02071
437. WO 91/13980
438. WO 91/13991
439. WO 91/15585
440. WO 93/01281
441. WO 93/01294
442. WO 93/18168
443. WO 93/21334
444. WO 94/00583
445. WO 95/19443
446. WO 97/17450
447. WO 97/17451
448. WO 97/41228
449. WO 97/48814
450. WO 98/01572
451. WO 98/02536
452. WO 98/02565
453. WO 98/45456
454. WO 98/45461
455. WO 99/32619
456. WO 99/53050
457. WO 00/44914
458. WO 00/49035
459. WO 00/63364
460. WO 00/68374
461. WO 99/53050
462. Wu *et al.* (1997) Methods in Gene Biotechnology (CRC Press, Inc.)
463. Xiaohui Wang *et al.* (2001) Gene 272(1-2): 249-255
464. Yamada *et al.* (1985) Proc Natl Acad Sci USA 82:6522-6526
465. Yarnell & Roberts (1999) Science 284:611-615
466. Yonaha & Proudfoot (1999) Molecular Cell 3:593-600
467. Yonaha & Proudfoot (2000) EMBO J. 19:3770-3777
468. Zhang *et al.* (2000) J Biol Chem 275(43):33850-33860
469. Zhao *et al.* (1999) Microbiol Mol Biol Rev 63:405-445
470. Zheng *et al.* (1997) Gene 186:55
471. Zhou *et al.* (1990) Mol. Cell. Biol. 10:4529
472. Zubko *et al.* (2000) Nat Biotechnol 18:442-445
473. Zupan *et al.* (2000) Plant J 23(1):11-28

# DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 109

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

# JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 109

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

**What is claimed is:**

1. A method for identification and/or isolation of transcription termination sequences for comprising the steps of:
   i) providing a screening construct or screening vector comprising
      a) a promoter sequence, and
      b) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences, and
      c) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic,

   wherein insertion of an efficient transcription terminator into said insertion site changes expression of said additional sequences by said promoter sequence in comparison to no insertion, and

   ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and

   iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector, and

   iv) introducing said screening construct or screening vector with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and

   v) identifying and/or selecting screening construct or screening vector with a changed readily detectable characteristic in comparison to no insertion, and

   vi) isolating the inserted DNA sequences from said identified and/or selected screening construct or screening vector for use as transcription termination sequences and – optionally – determining their sequence.

2. The method of Claim 1, wherein said insertion site is localized in relation to said additional sequences at a position selected from group of:
   i) upstream of the additional sequences between said promoter and said additional sequences, and
   ii) downstream of the additional sequences, and
   iii) in between said additional sequences.

3. The method of Claim 1 or 2, wherein said DNA sequences to be assessed for their transcription termination efficiency are inserted into said insertion site in form of an inverted repeat.

4. The method of any of Claim 1 to 3, wherein the method for identification and isolation of transcription termination sequences comprises the steps of:
   i) providing a screening construct or screening vector comprising in 5' to 3' direction
      a) a promoter sequence, and
      b) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic, and
      c) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences,
   ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and
   iii) inserting at least two copies of a specific DNA sequence of said DNA sequences in form of an inverted repeat into said insertion site of said screening construct or screening vector, wherein insertion of an inverted repeat of an efficient transcription terminator into said insertion site allows expression of said additional sequences by said promoter sequence in comparison to no insertion, and
   iv) introducing said screening constructs or screening vectors with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and
   v) identifying and/or selecting screening constructs or screening vectors with said readily detectable characteristic in comparison to no insertion, and
   vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

5. The method of Claim 1 to 2, wherein the method for identification and isolation of transcription termination sequences comprises the steps of:
   i) providing a screening construct or screening vector comprising in 5' to 3' direction
      a) a promoter sequence, and
      b) one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences, and
      c) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic,

   wherein insertion of an efficient transcription terminator into said insertion site suppresses expression of said additional sequences by said promoter sequence in comparison to no insertion, and
   ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and
   iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector , and
   iv) introducing said screening construct or screening vector with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and

v) identifying and/or selecting screening constructs or screening vectors with a changed readily detectable characteristic in comparison to no insertion, and
vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

6. The method of Claim 5, wherein said additional sequences are an inverted repeat of a known transcription terminator sequences localized in a way said the second copy of said repeat downstream from said promoter sequence is in the orientation to constitute a functional transcription terminator, and wherein said screening construct or screening vector is comprising further sequences between said insertion site and said promoter which result in an readily detectable phenotype, and wherein absence of an efficient transcription terminator in said insertion site will lead to silencing of expression of said further sequences in consequence of expression of said inverted repeat of said known transcription terminator.

7. The method of Claim 1 or 2 , wherein the method for identification and isolation of transcription termination sequences comprises the steps of:
i) providing a screening construct or screening vector comprising in 5' to 3' direction
a) a promoter sequence, and
b) at least one additional sequence which causes upon expression under said promoter sequence a readily detectable characteristic, and embedded into said additional sequences one or more insertion sites – preferably a restriction or recombination site - for insertion of DNA sequences,
wherein insertion of an efficient transcription terminator into said insertion site suppresses full-length transcription of said additional sequences by said promoter sequence in comparison to no insertion, and
ii) providing one or more DNA sequences to be assessed for their transcription termination capability, and
iii) inserting one or more copies of said DNA sequences into said insertion site of said screening construct or screening vector , and
iv) introducing said screening constructs or screening vectors with said inserted DNA sequences into an *in vitro* or *in vivo* transcription system suitable to induce expression from said promoter sequence, and
v) identifying and/or selecting screening constructs or screening vectors with a changed readily detectable characteristic in comparison to no insertion, and
vi) isolating the inserted DNA sequences from said identified and/or selected screening constructs or screening vectors for use as transcription termination sequences and – optionally – determining their sequence.

8. The method of Claim 7, wherein said insertion site is localized within an intron comprised in said additional sequence.

9. The method of Claim 7, wherein said additional sequence comprises a first 5'-part which is encoding for a marker protein and second 3'-part which is not protein encoding and has no homology to sequences endogenous to said transcription system, and wherein said insertion site is localized between said 5'-part and 3'-part, and wherein for the purpose of selection and/or identification a second expression cassette is introduced into said transcription system comprising a promoter function in said transcription system and operably linked thereto a sequence encoding for an antisense or double-stranded RNA molecule complementary to said 3'-part.

10. The method of any of Claim 1 to 9, wherein said additional sequence is selected from the group consisting of positive selection marker, negative selection marker, counter selection marker, reporter genes, and toxic genes.

11. The method of Claim 108, wherein said toxic gene is a construct for gene silencing of an essential endogenous gene or a selectable marker.

12. The method of any of Claim 1 to 11, wherein said DNA sequence to be assessed for their transcription termination efficiency is provided by a method selected from the group consisting of:

i) provision of a selected sequence by amplification from a host genome, and

ii) provision of a library of sequences by fragmentation of a host genome, and

iii) provision of a library of synthetic sequences, and

iv) provision of selected sequence by a method as claimed in any of Claim 13 to 15.

13. The method of Claim 12, wherein said DNA sequence to be assessed for their transcription termination efficiency is derived from an eukaryotic organism.

14. The method of Claim 12 or 13, wherein said DNA sequence to be assessed for their transcription termination efficiency is derived from a plant.

15. The method of any of Claim 1 to 14, wherein said DNA sequences to be assessed for their transcription termination efficiency are inserted into said insertion site by a method selected from the group consisting of:

i) recombinational cloning, and

ii) insertion by sequence specific restriction and ligation.

16. A method for identification and/or isolation of intergenic regions with transcription termination potential said method including at least the steps of

a) identification and/or isolation or isolation of intergenic regions between paired genes having an intergenic distance of about 400 to 3,000 base pairs, and

b) identification and/or isolation of intergenic sequences which are flanked on both sides by genes having a high expression level.

17. The method of Claim 16 further comprising the steps of
    c) identification and/or isolation of intergenic sequences which are flanked on both sides by genes having an expression pattern which is preferably independent from the expression pattern of the other paired gene, and
    d) identification and/or isolation of intergenic sequences which are flanked on one or – preferably - both sides by genes having a low variability in length of the mRNA transcript derived from said paired genes.

18. The method of Claim 16 or 17, wherein said intergenic region
    a) is localized between genes which have a tail-to-tail localization and / or
    b) has a length measured from the respective stop codons of the flanking genes from about 700 to about 2000 base pairs.

19. A transgenic expression construct comprising in 5'-3'-direction
    a) a promoter sequence functional in plants, and
    b) a nucleic acid sequence of interest of to be expressed operably linked to said promoter a), and
    c) at least one sequence selected from the group consisting of
        i) the sequences described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, and 108, and
        ii) the sequences having a homology of at least 60%, preferably 80%, more preferably 90%, most preferably 95% with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and
        iii) the sequences hybridizing under low stringency conditions, preferably under high stringency conditions with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and
        iv) a fragment of at least 50 consecutive base pairs, preferably at least 100 consecutive base pairs, more preferably at least 250 consecutive base pairs, most preferably at least 500 consecutive base pairs of a sequence described under i), ii), and iii),
    wherein said sequence c) is heterolog with respect to said promoter a) and/or said nucleic acid of interest b) and is mediating termination of expression of induced from said promoter a).

20. A transgenic expression construct comprising at least two expression cassettes having a structure comprising in 5'-3'-direction

a1) a first promoter sequence functional in plants, and

b1) a first nucleic acid sequence of interest of to be expressed operably linked to said promoter a1), and

c) at least one sequence selected from the group consisting of

i) the sequences described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, and 108, and

ii) the sequences having a homology of at least 60%, preferably 80%, more preferably 90%, most preferably 95% with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and

iii) the sequences hybridizing under low stringency conditions, preferably under high stringency conditions with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and

iv) a fragment of at least 50 consecutive base pairs, preferably at least 100 consecutive base pairs, more preferably at least 250 consecutive base pairs, most preferably at least 500 consecutive base pairs of a sequence described under i), ii), and iii),

and,

b2) a second nucleic acid sequence of interest of to be expressed, and

a2) a second promoter sequence functional in plants operably linked to said nucleic acid sequence of interest b2),

wherein said sequence c) is heterolog with respect to at least one element selected from promoter a1), promoter a2), nucleic acid of interest b1) and nucleic acid of interest b2), and is mediating termination of expression of induced from said promoters a1) and a2).

21. The use of a sequence selected from the group consisting of

i) the sequences described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, and 108, and

ii) the sequences having a homology of at least 60%, preferably 80%, more preferably 90%, most preferably 95% with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42,

43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and

iii) the sequences hybridizing under low stringency conditions, preferably under high stringency conditions with a sequences described by described by SEQ ID NO: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 92, 93, 107, or 108, capable to terminate transcription in a plant cell or organism, and

iv) a fragment of at least 50 consecutive base pairs, preferably at least 100 consecutive base pairs, more preferably at least 250 consecutive base pairs, most preferably at least 500 consecutive base pairs of a sequence described under i), ii), and iii).

as transcription terminator and or isolator.

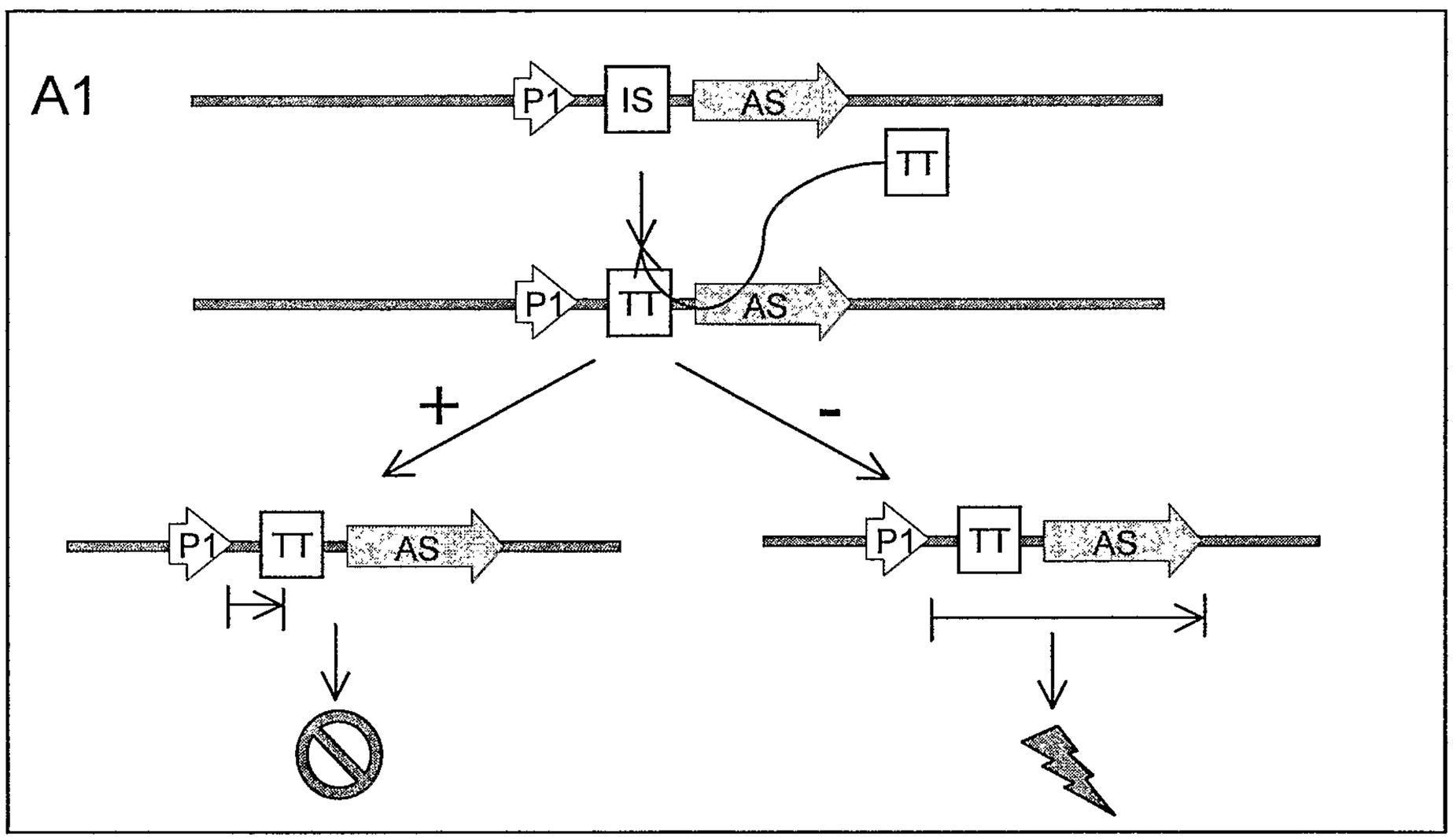

A1
P1
IS
AS
TT
P1
TT
AS
+
-
P1
TT
AS
P1
TT
AS
A2
P1
M1
IS
⊥
T
P2
M2
T
TT
P1
M1
TT
⊥
T
P2
M2
+
-
P1
M1
TT
⊥
T
P2
M2
T
P1
M1
TT
⊥
T
P2
M2
T
GS

Fig. 1

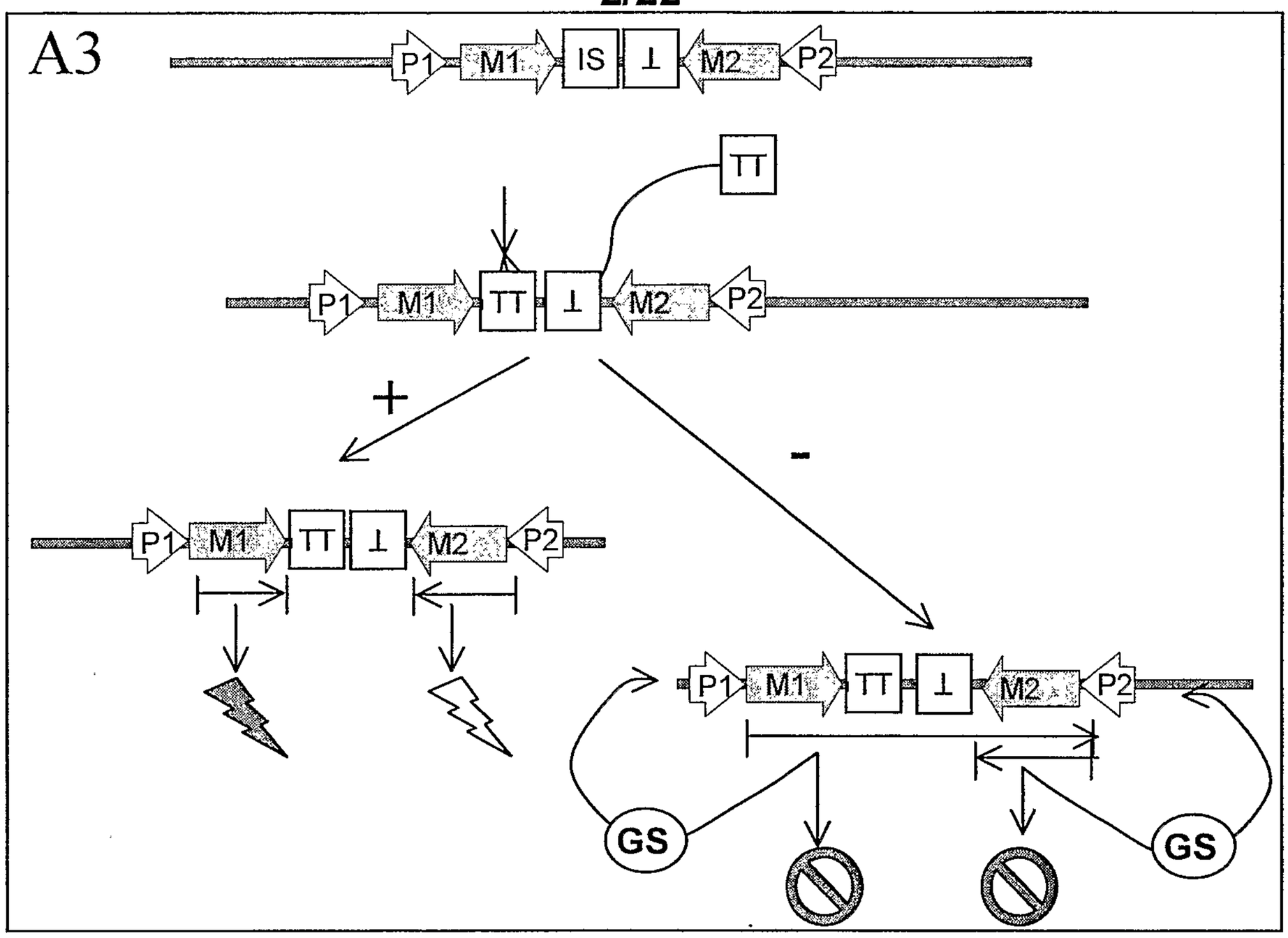
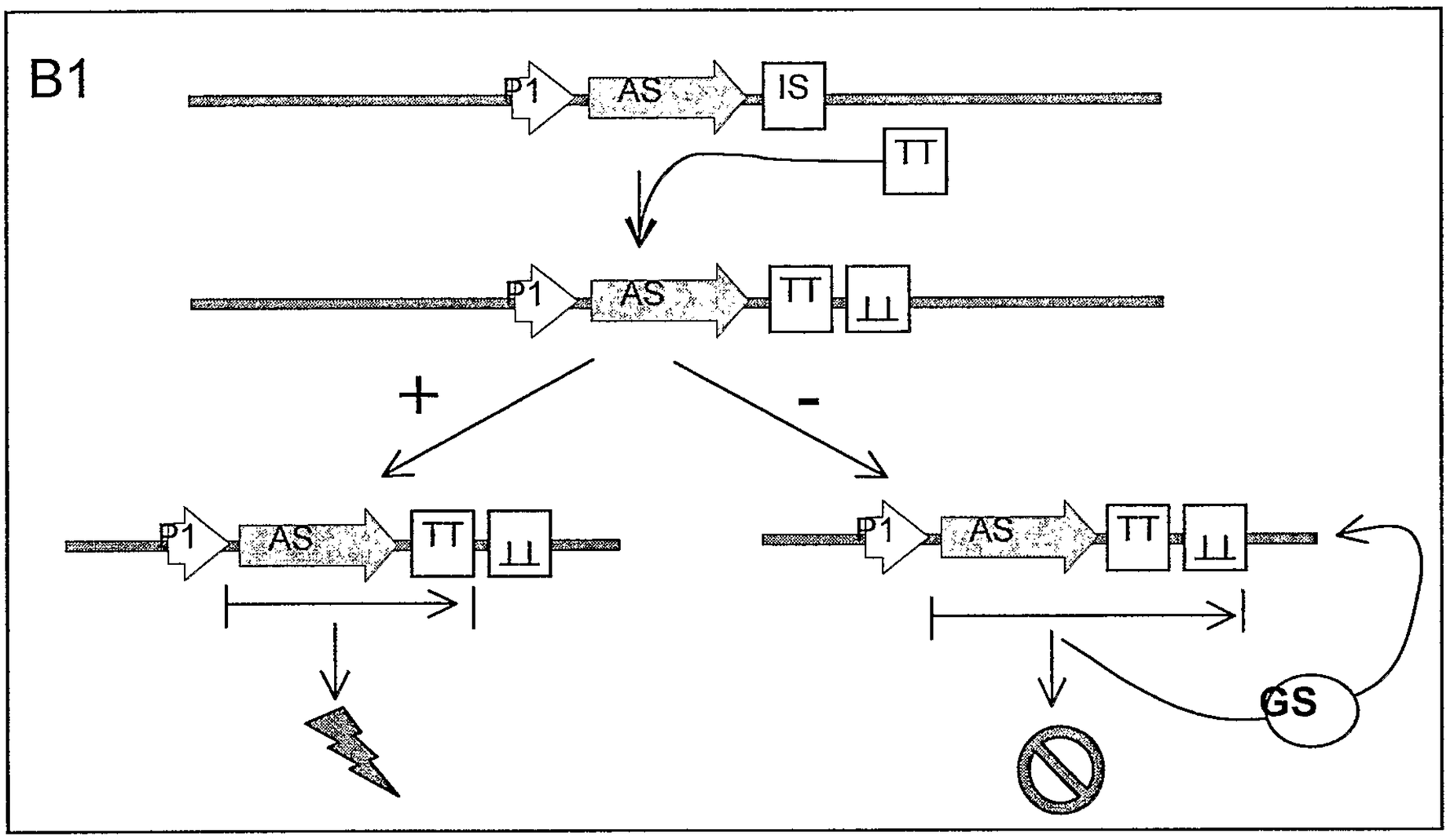



Fig. 1-B


B1
P1
AS
IS
TT
P1
AS
TT
⊥⊥
+
-
P1
AS
TT
⊥⊥
P1
AS
TT
⊥⊥
GS

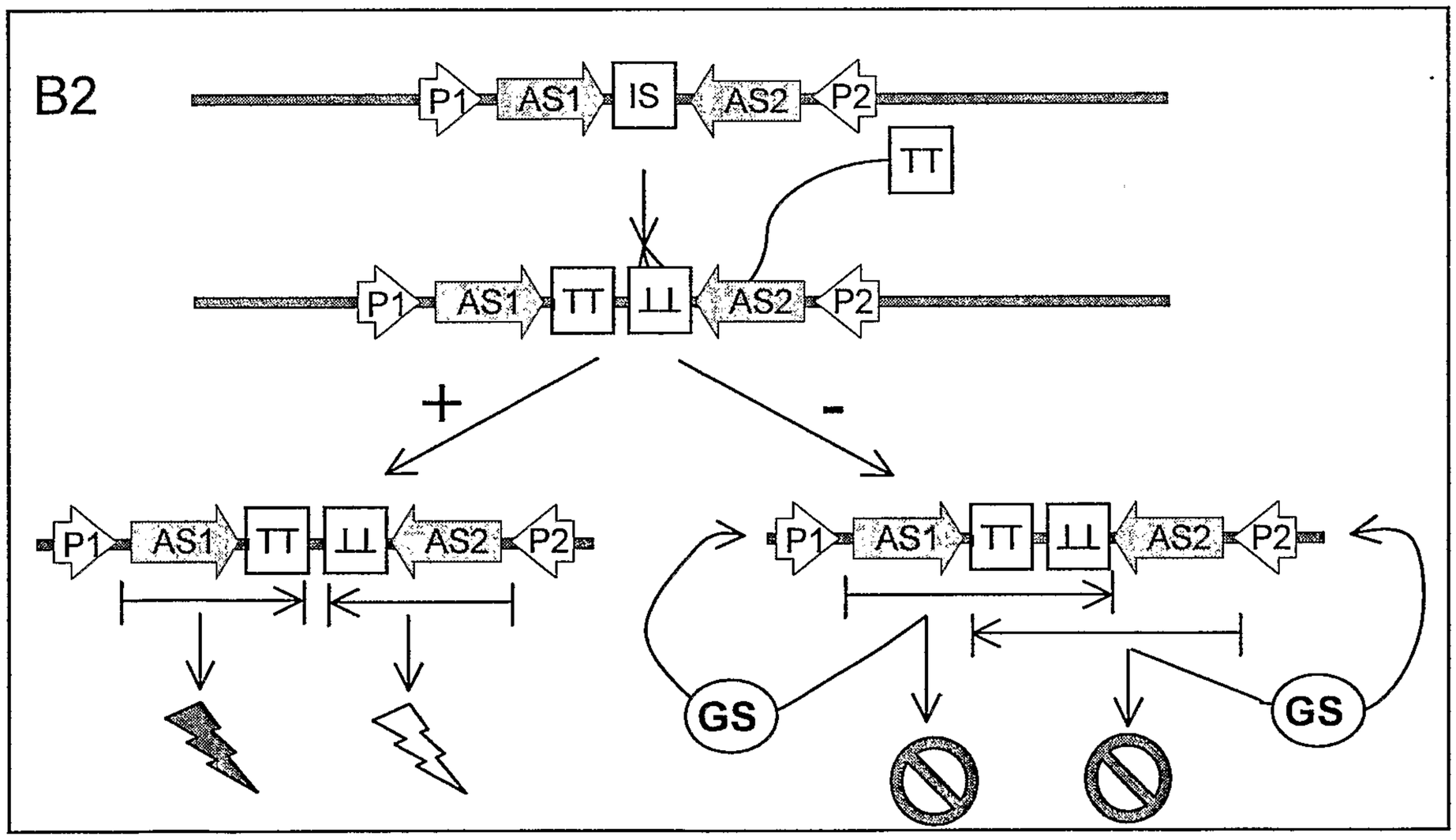

B2
P1
AS1
IS
AS2
P2
TT
P1
AS1
TT
⊥⊥
AS2
P2
+
-
P1
AS1
TT
⊥⊥
AS2
P2
P1
AS1
TT
⊥⊥
AS2
P2
GS
GS

Fig. 2

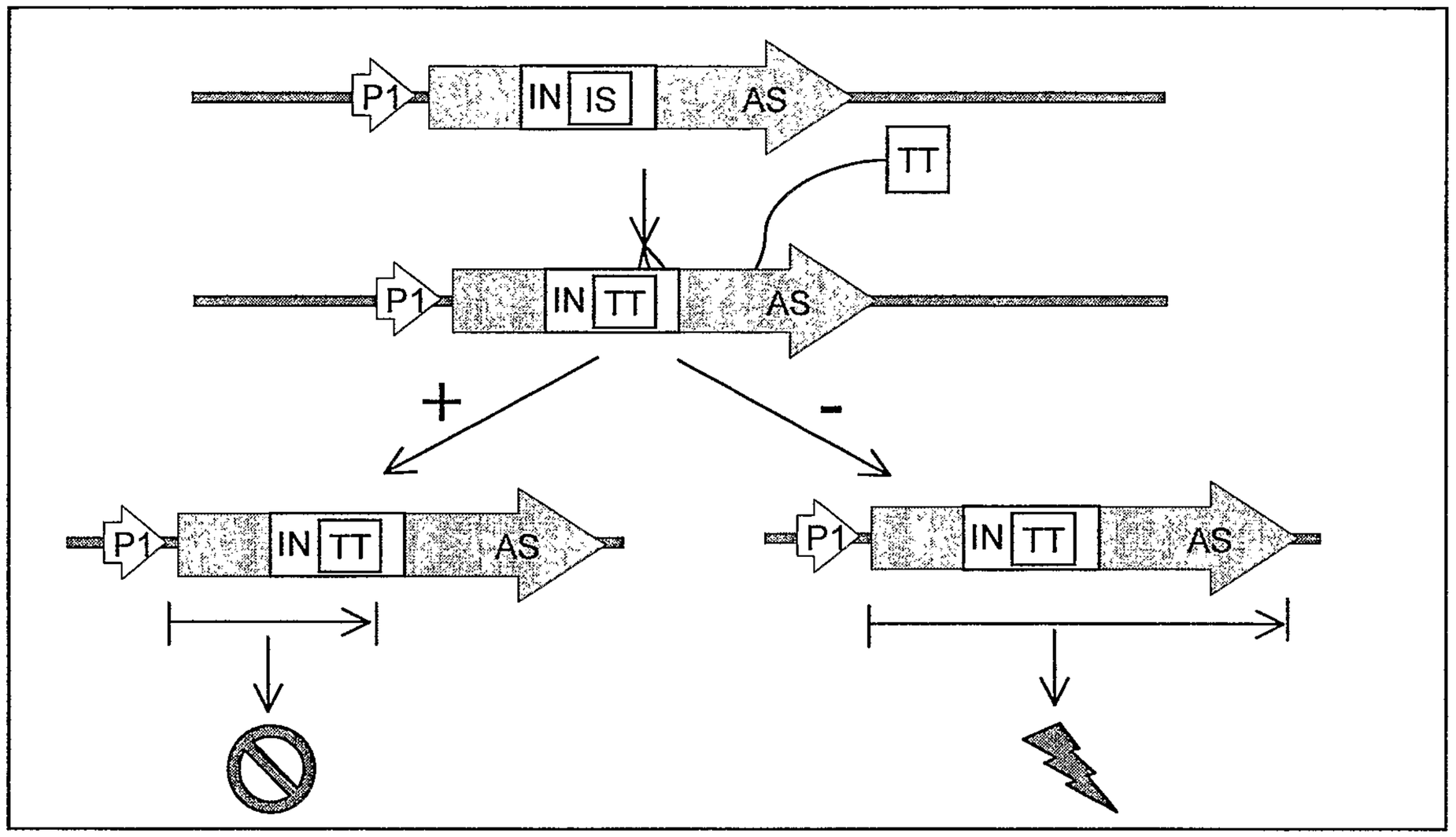



Fig. 3

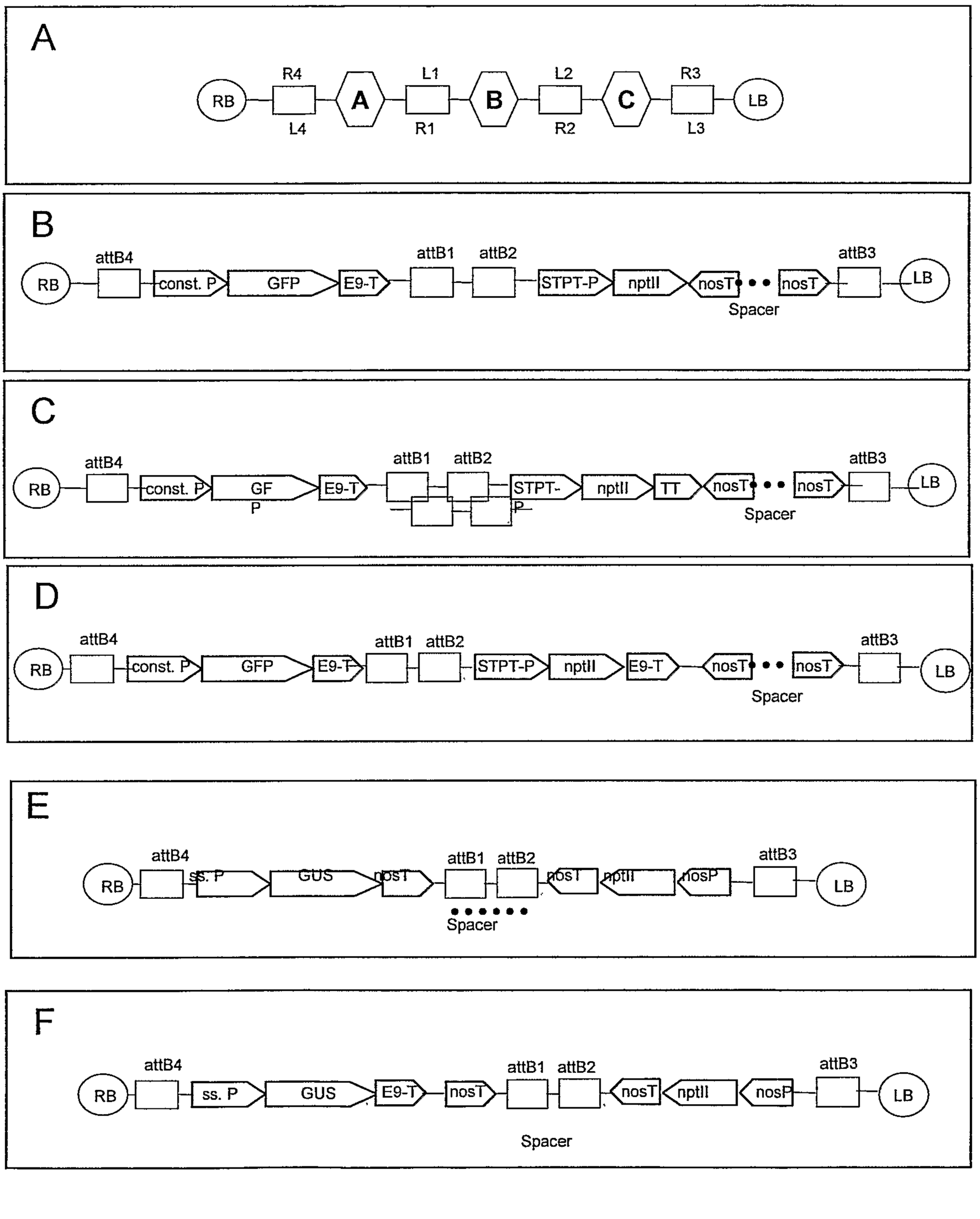
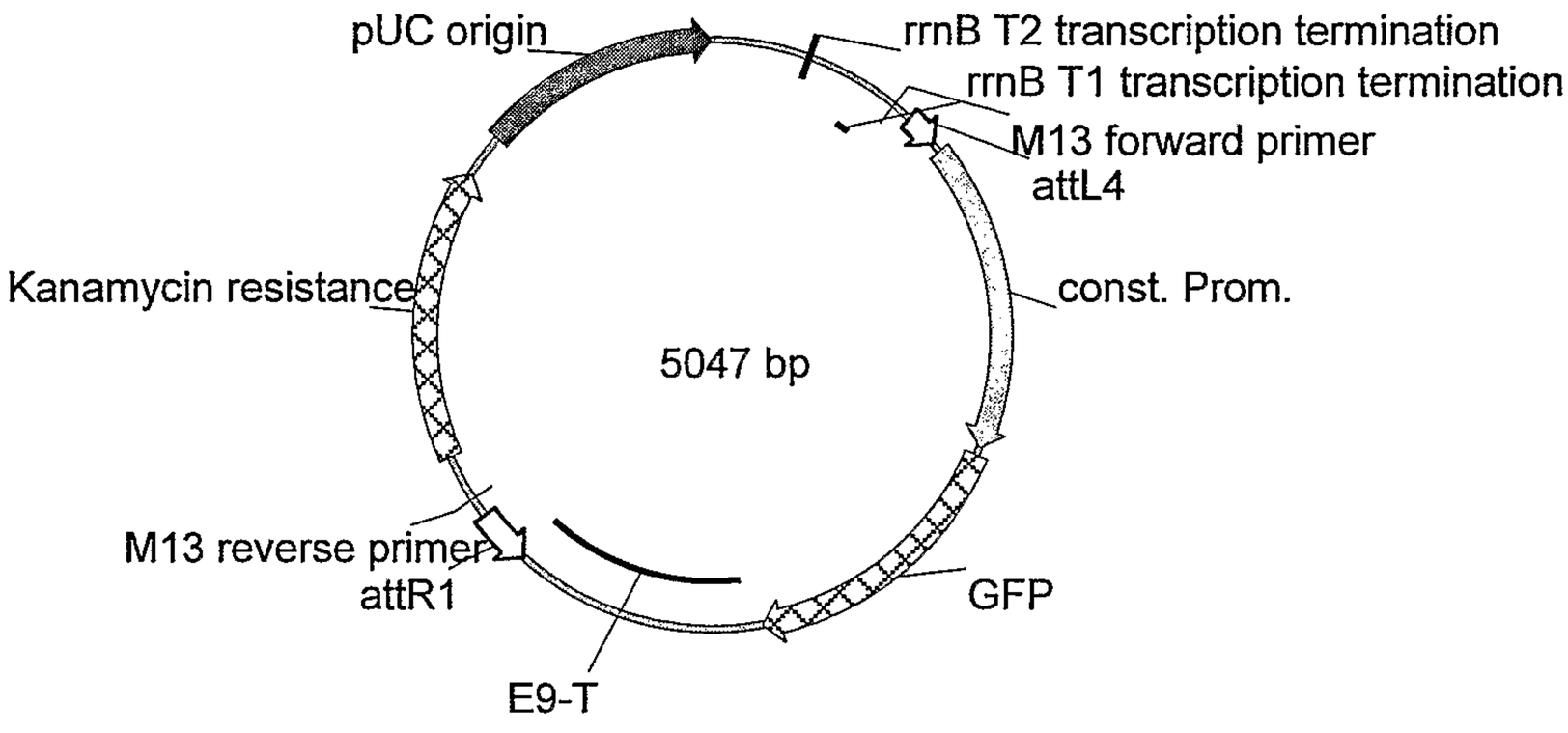
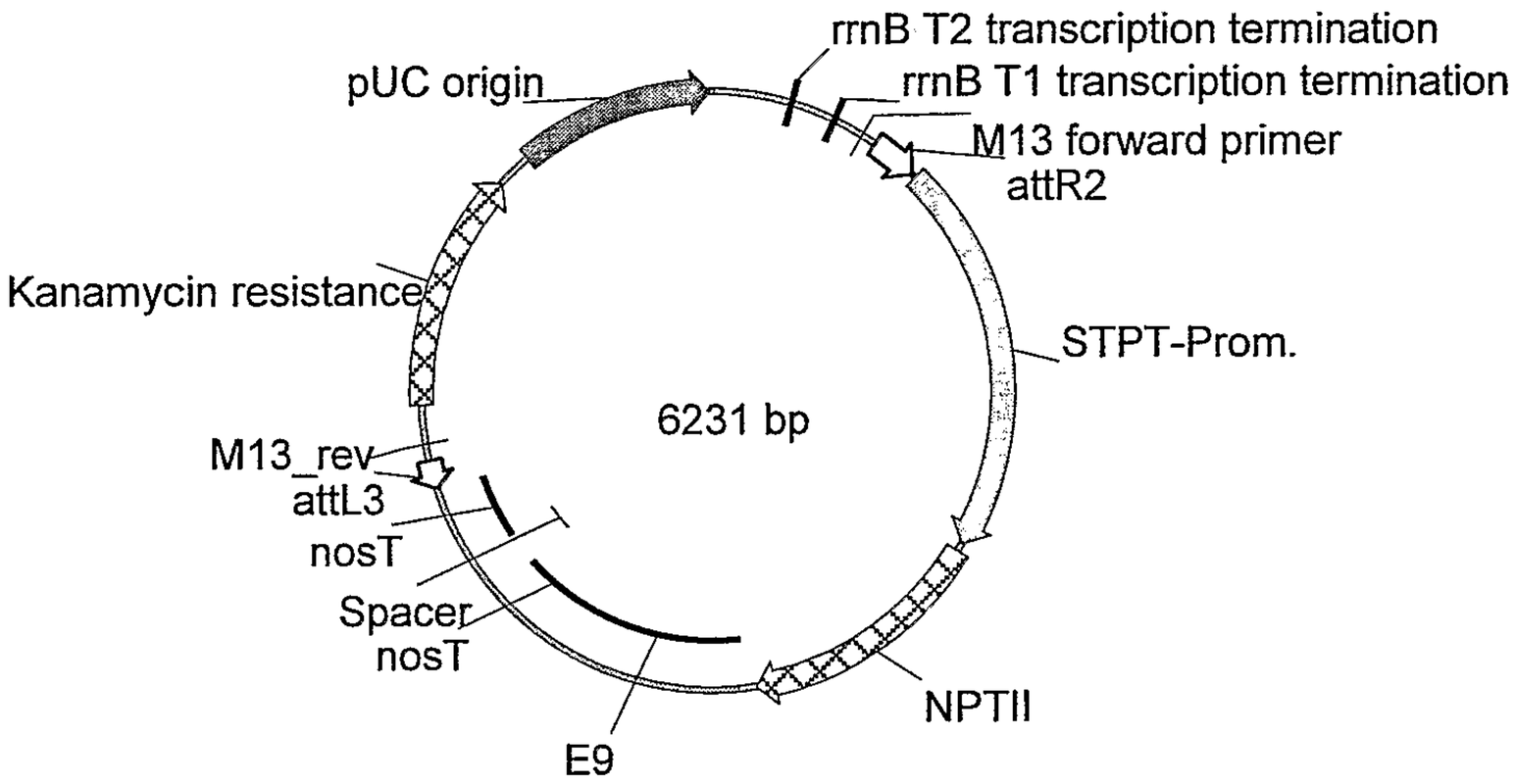

A
RB
R4
L4
A
L1
R1
B
L2
R2
C
R3
L3
LB
B
RB
attB4
const. P
GFP
E9-T
attB1
attB2
STPT-P
nptII
nosT
nosT
attB3
LB
Spacer
C
RB
attB4
const. P
GF
P
E9-T
attB1
attB2
STPT-
P
nptII
TT
nosT
nosT
attB3
LB
Spacer
D
RB
attB4
const. P
GFP
E9-T
attB1
attB2
STPT-P
nptII
E9-T
nosT
nosT
attB3
LB
Spacer
E
RB
attB4
ss. P
GUS
nosT
attB1
attB2
nosT
nptII
nosP
attB3
LB
Spacer
F
RB
attB4
ss. P
GUS
E9-T
nosT
attB1
attB2
nosT
nptII
nosP
attB3
LB
Spacer

Fig. 4

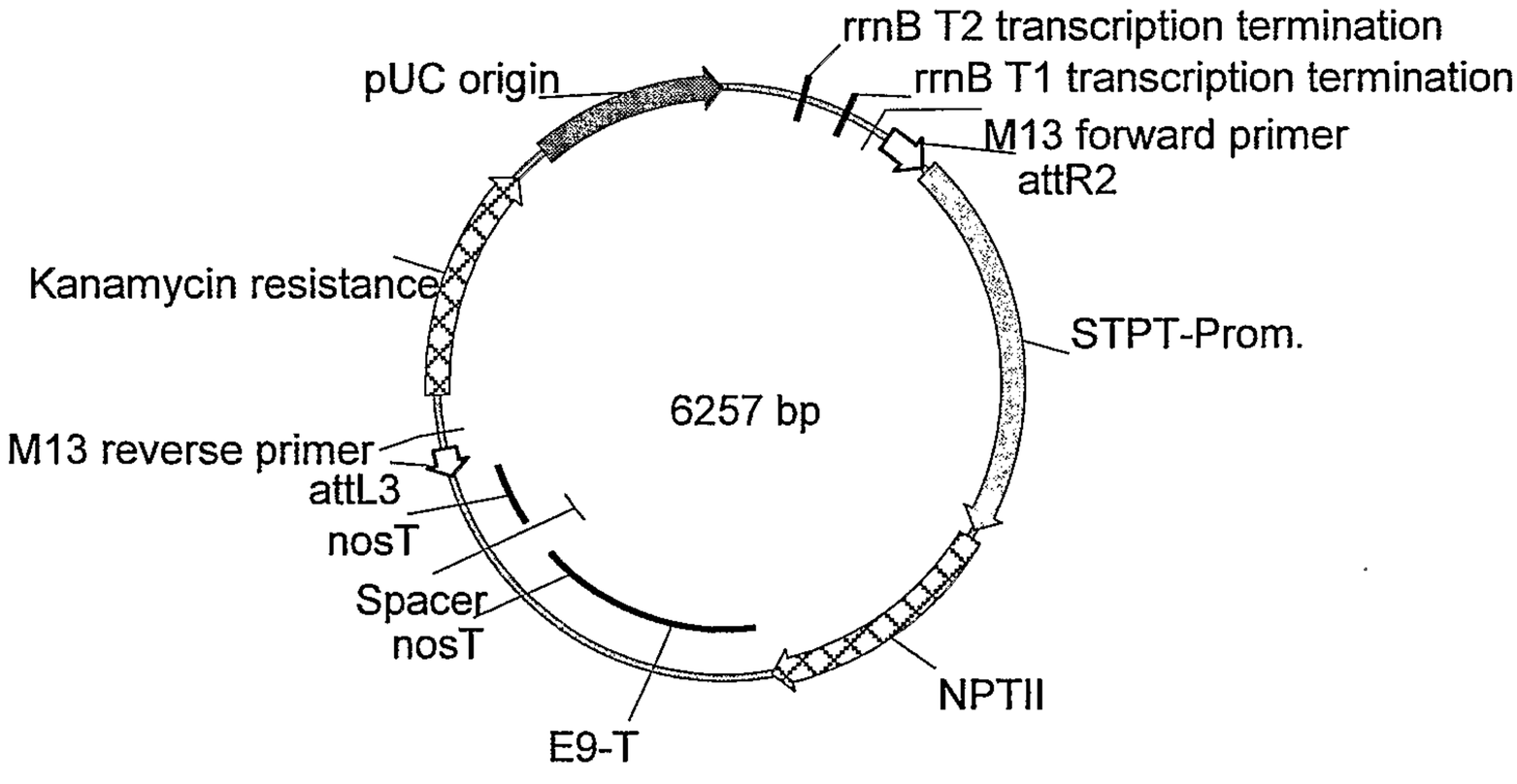



Lo484-pENTR-A1-inv-35s-GFP-E9

Fig. 5

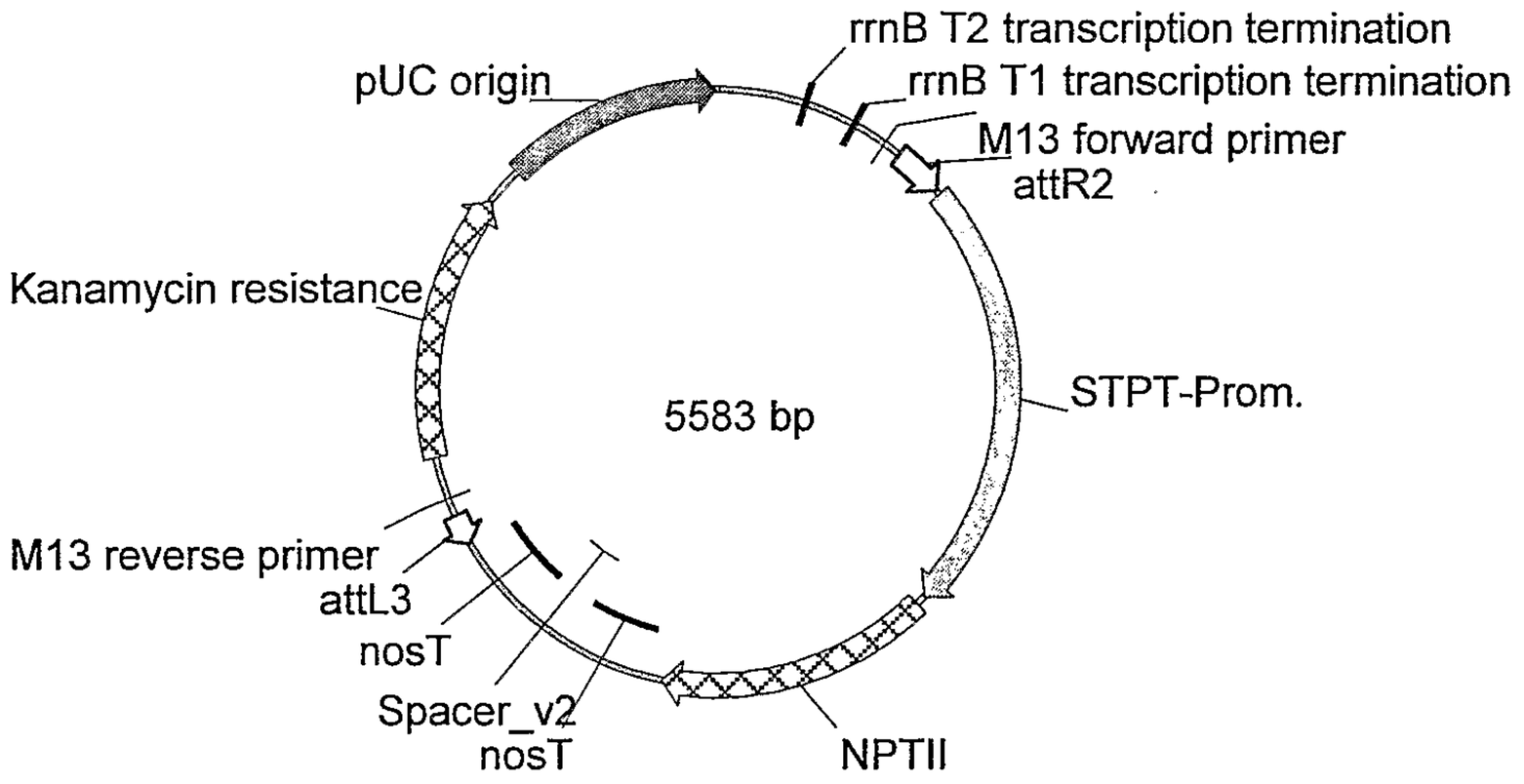



**Lo522a-pENTR-C1-R2-STPT-nptII-E9-lRnos-L3**

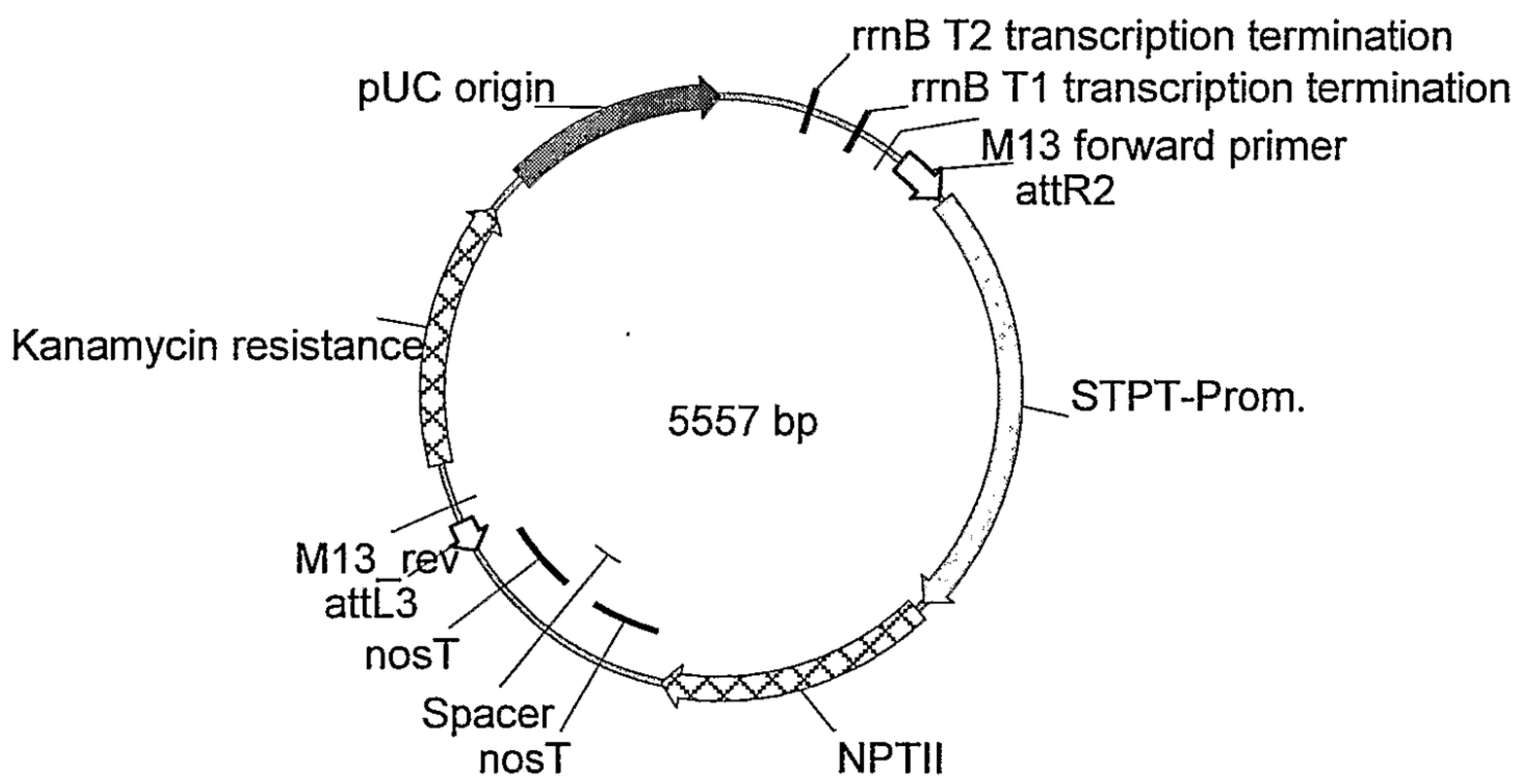



**Lo522b-pENTR-C1-R2-STPT-nptII-E9-lRnos-L3**

Fig. 6

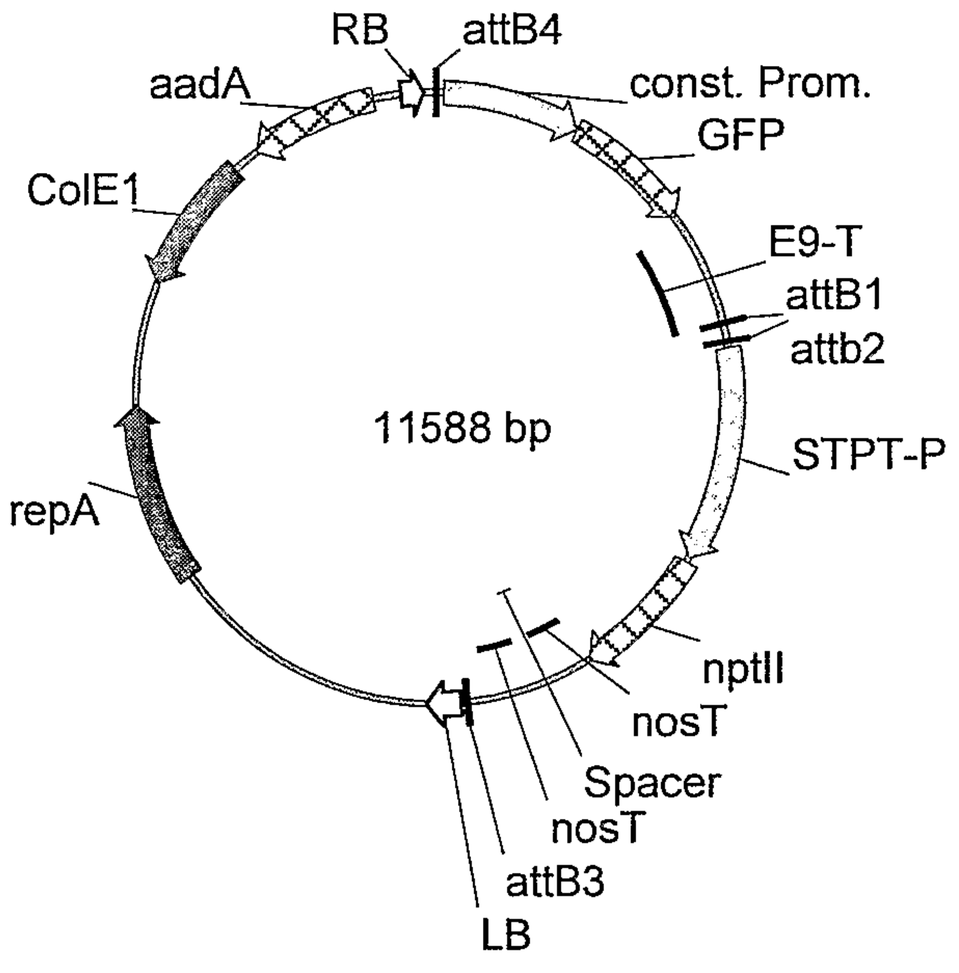



**Lo503b-pENTR-C1-R2-STPT-nptII-IRnos-L3**

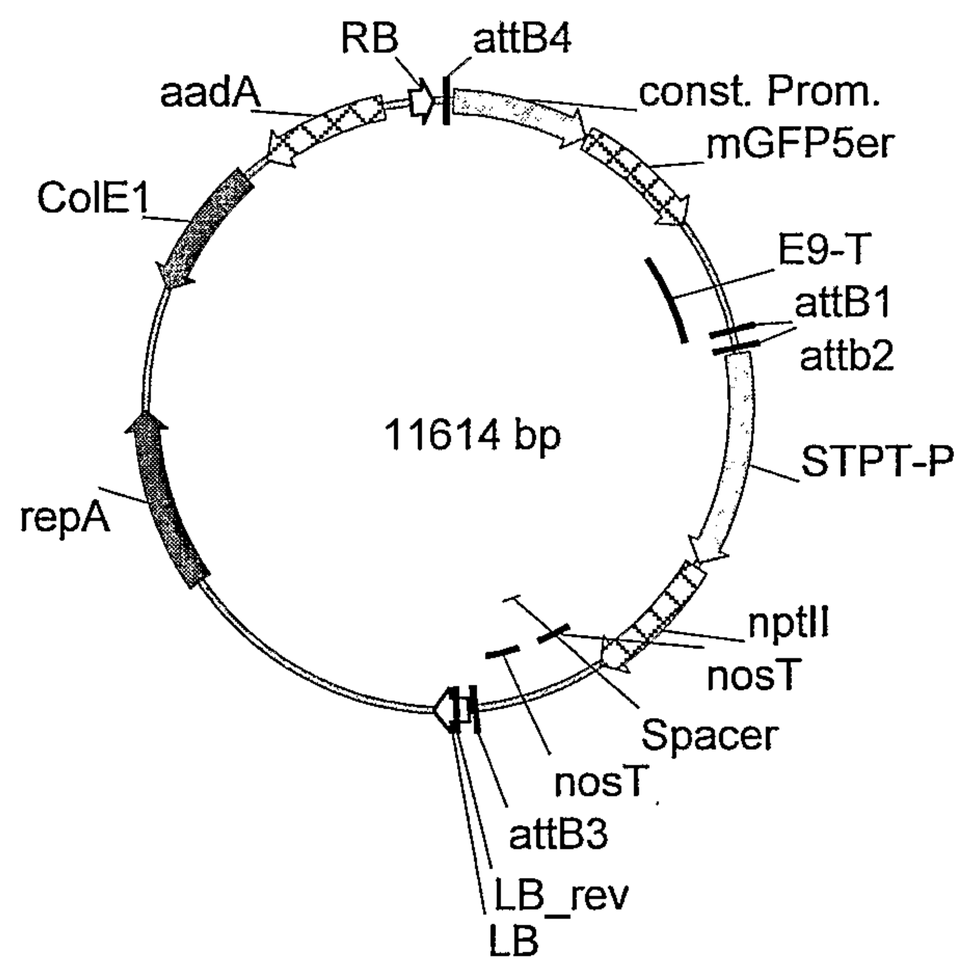



**Lo503a-pENTR-C1-R2-STPT-nptII-IRnos-L3**

Fig. 7

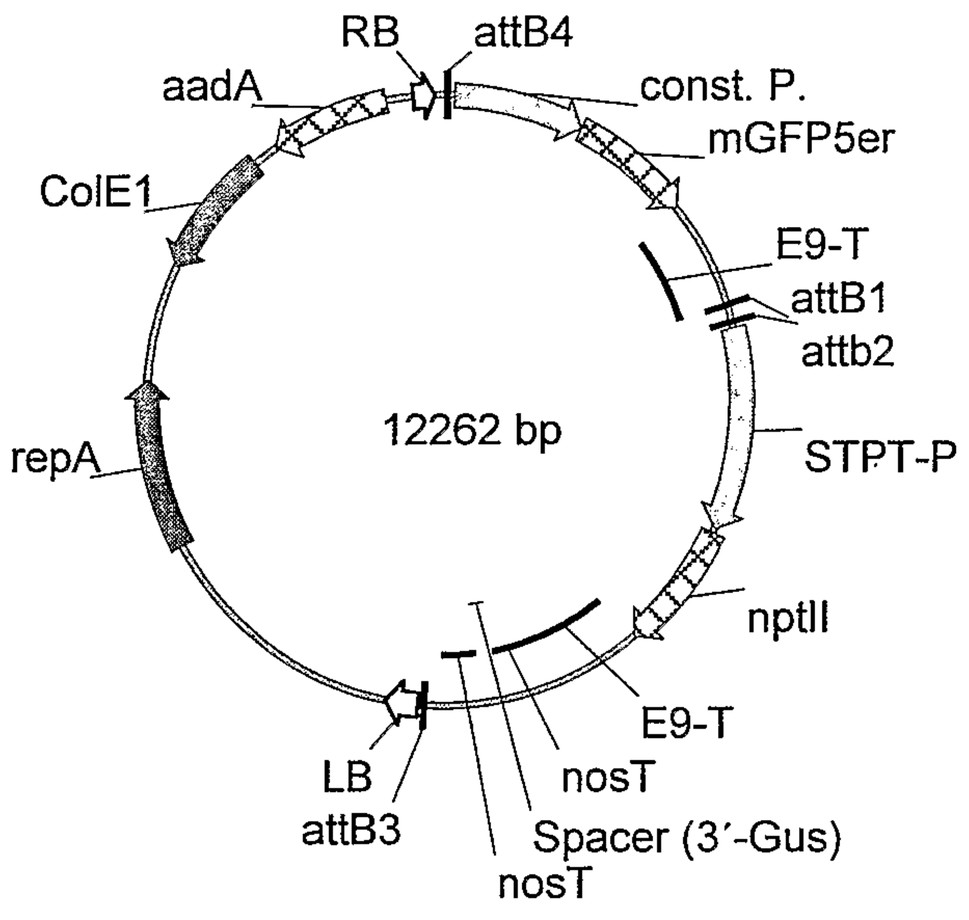



**Lo523a-pSun1-R4-Lo484::Lo376::Lo503a-R3**

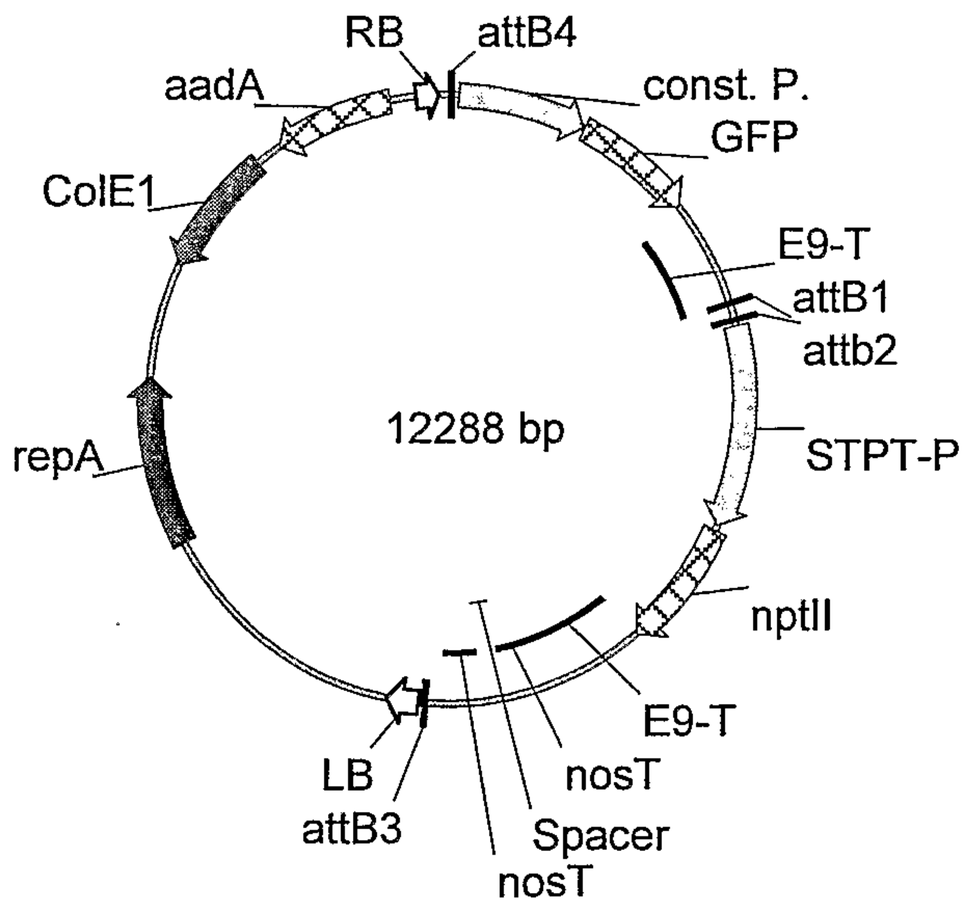



**Lo523b-pSun1-R4-Lo484::Lo376::Lo503b-R3**

Fig. 8

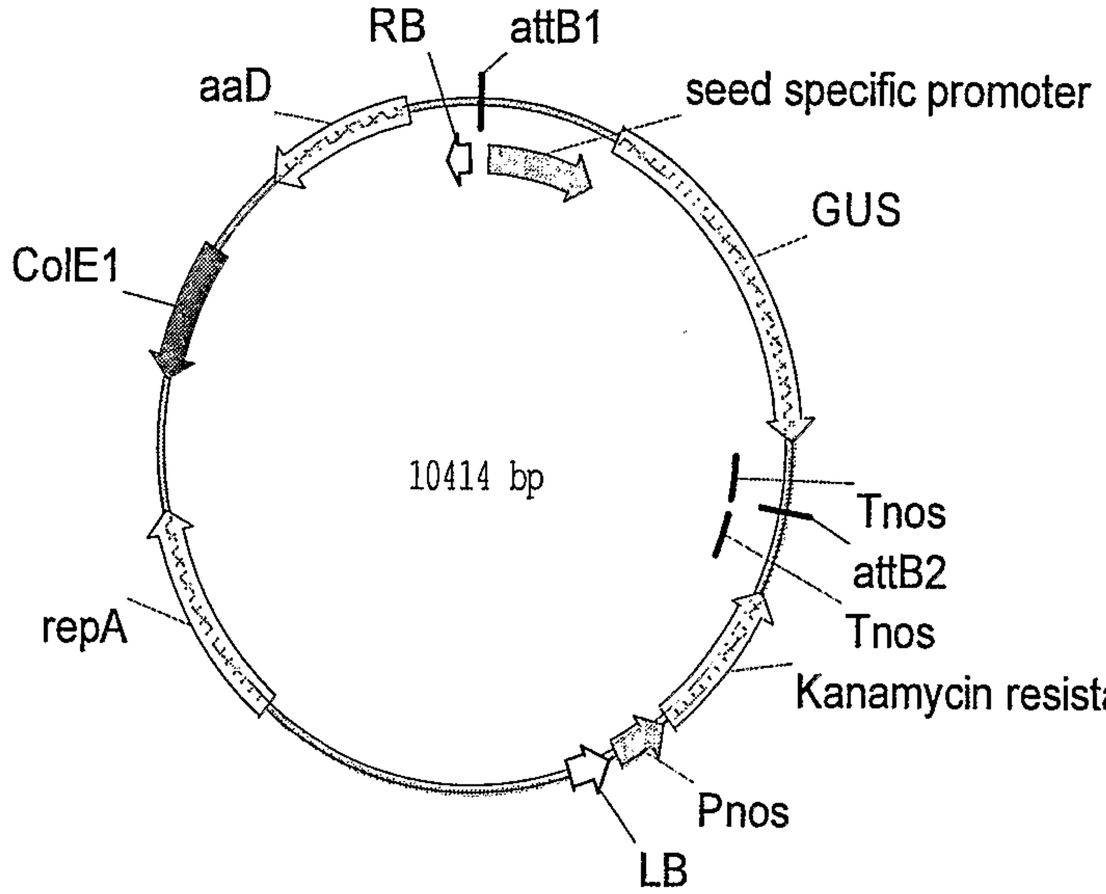



**Lo546a-pSun1-R4-Lo484::Lo376::Lo522a-R3**




**Lo546b-pSun1-R4-Lo484::Lo376::Lo522b-R3**

Fig. 9



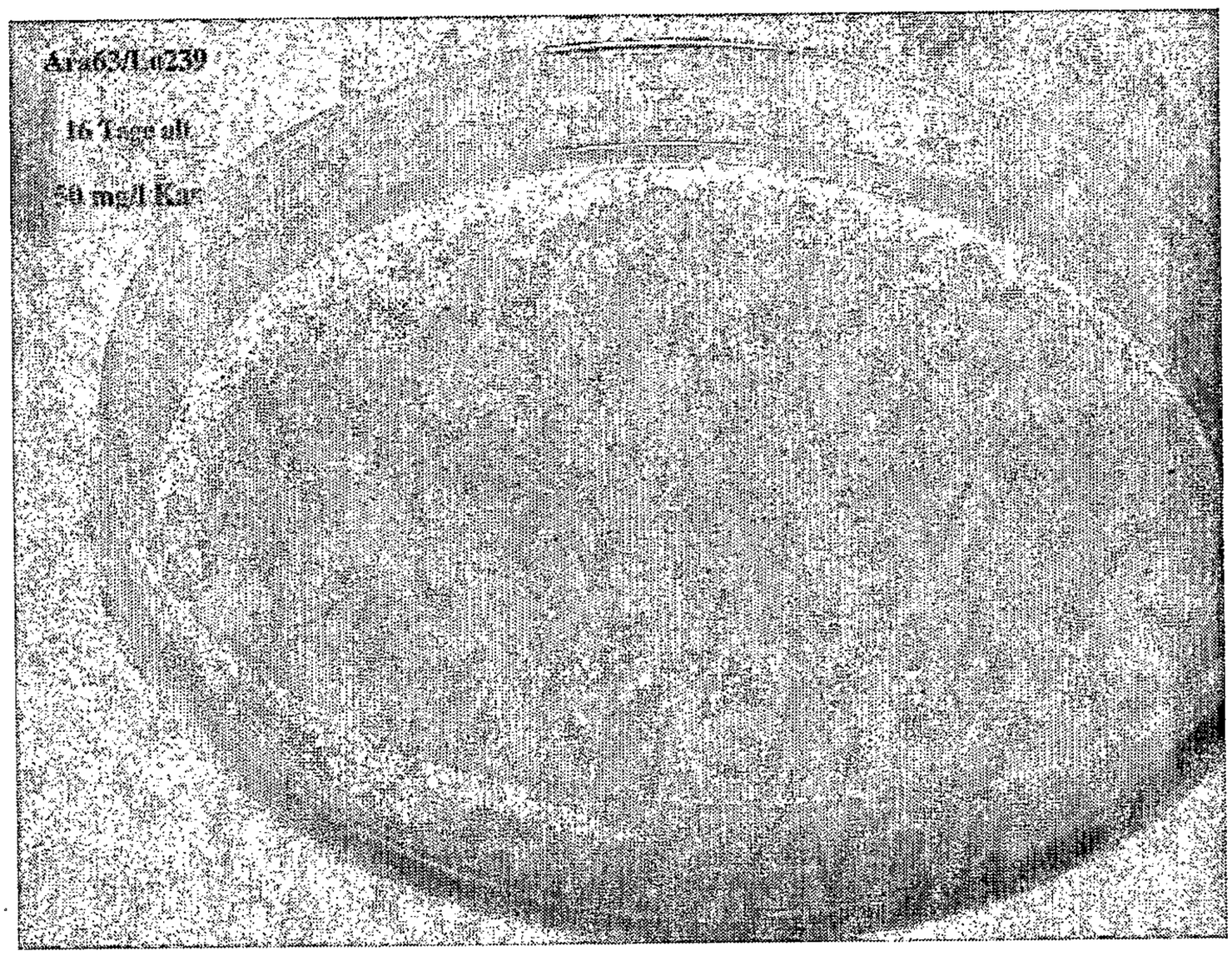

Lo239-pSUN3-GWs-B1-BnAK700::GUS::nosT-B2




Lo657 5-pSun3-GWs-B1-BnAK700::GUS::E9::nosT-B2

Fig.10

A


Ara63/Lo239
16 Tage alt
50 mg/l Kan

B


Ara176/Lo657
16 Tage alt
50 mg/l Kan

Fig. 11

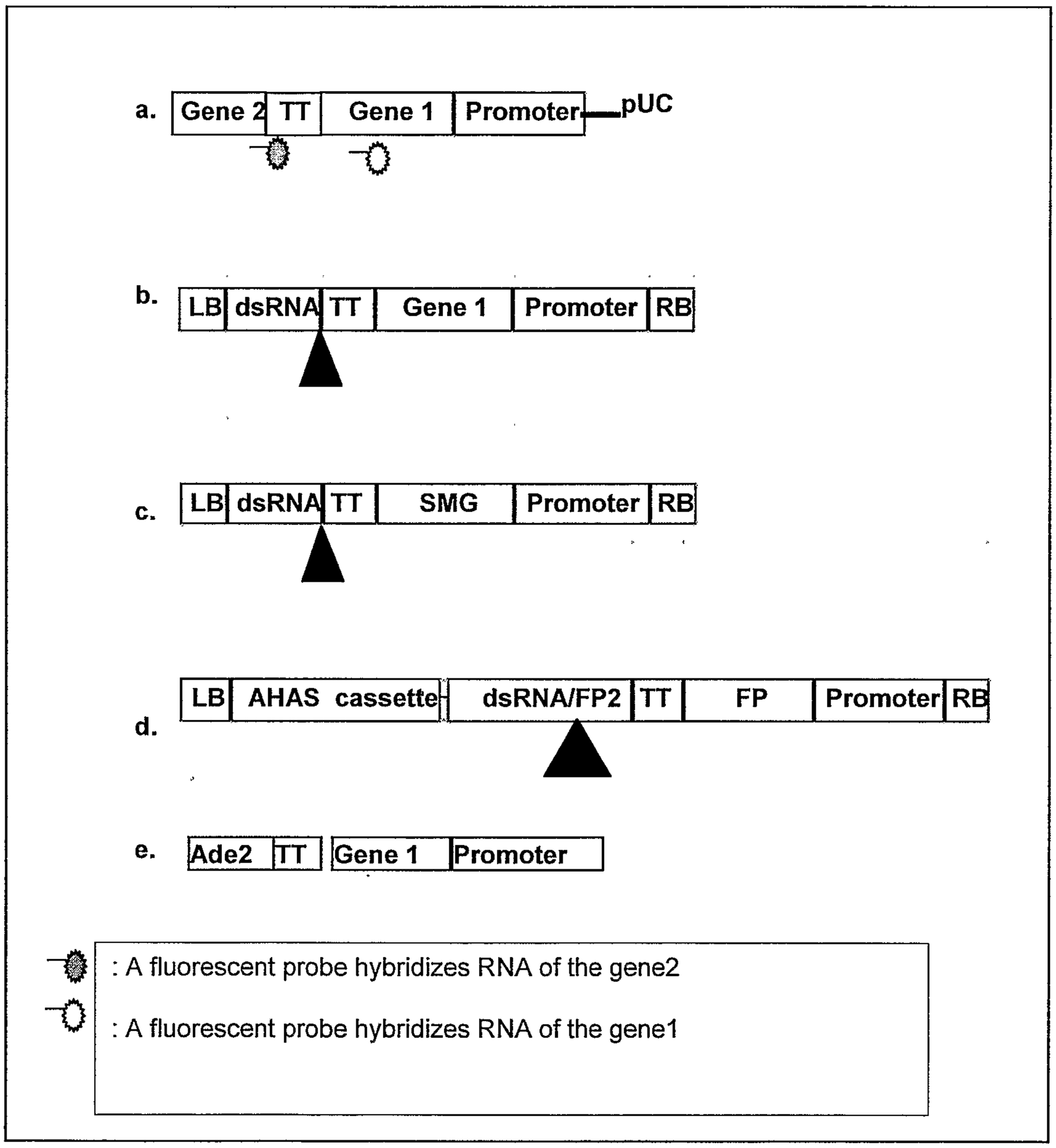

a.
Gene 2
TT
Gene 1
Promoter
pUC
b.
LB
dsRNA
TT
Gene 1
Promoter
RB
c.
LB
dsRNA
TT
SMG
Promoter
RB
d.
LB
AHAS cassette
dsRNA/FP2
TT
FP
Promoter
RB
e.
Ade2
TT
Gene 1
Promoter
: A fluorescent probe hybridizes RNA of the gene2
: A fluorescent probe hybridizes RNA of the gene1

Fig.12

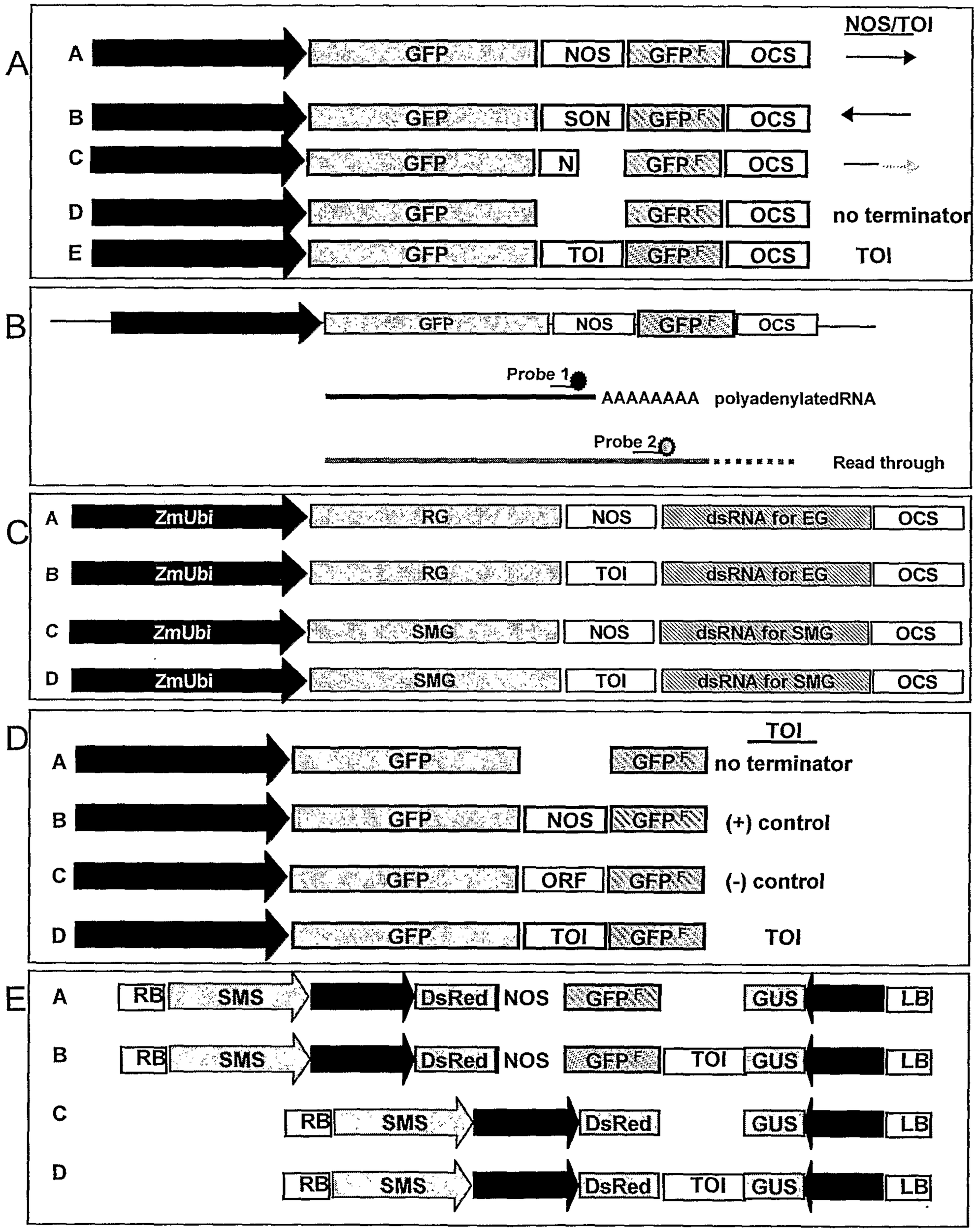
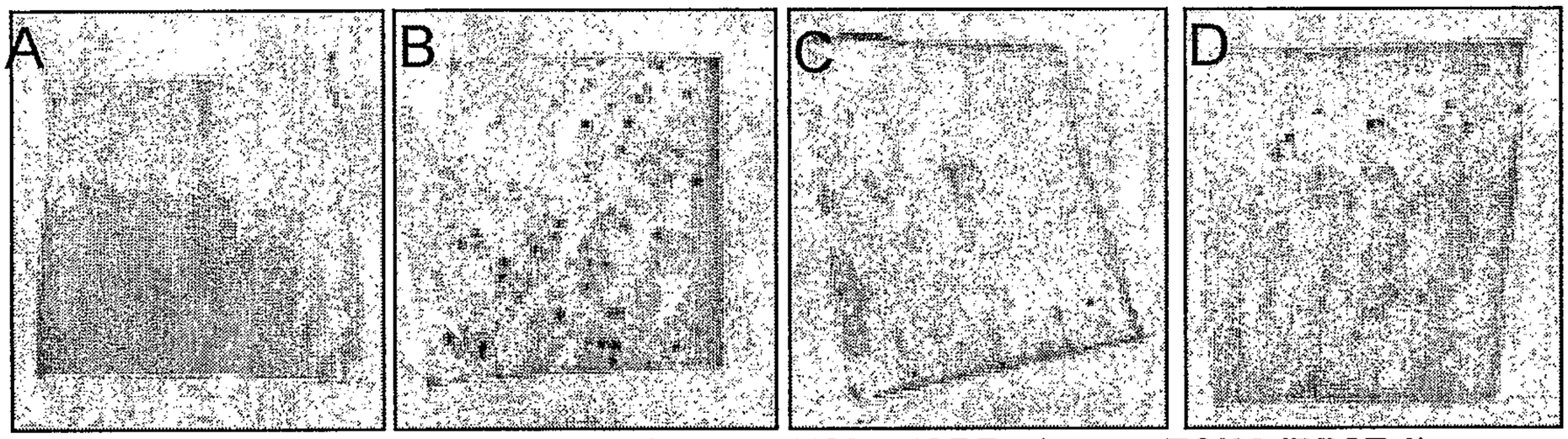

A
NOS/TOI
A
GFP
NOS
GFP F
OCS
B
GFP
SON
GFP F
OCS
C
GFP
N
GFP F
OCS
D
GFP
GFP F
OCS
no terminator
E
GFP
TOI
GFP F
OCS
TOI
B
GFP
NOS
GFP F
OCS
Probe 1
AAAAAAAA polyadenylatedRNA
Probe 2
Read through
C
A
ZmUbi
RG
NOS
dsRNA for EG
OCS
B
ZmUbi
RG
TOI
dsRNA for EG
OCS
C
ZmUbi
SMG
NOS
dsRNA for SMG
OCS
D
ZmUbi
SMG
TOI
dsRNA for SMG
OCS
D
TOI
A
GFP
GFP F
no terminator
B
GFP
NOS
GFP F
(+) control
C
GFP
ORF
GFP F
(-) control
D
GFP
TOI
GFP F
TOI
E
A
RB
SMS
DsRed
NOS
GFP F
GUS
LB
B
RB
SMS
DsRed
NOS
GFP F
TOI
GUS
LB
C
RB
SMS
DsRed
GUS
LB
D
RB
SMS
DsRed
TOI
GUS
LB

Fig. 13


A
B
C
D
pTOI04 (null)
pTOI05 (NOS,+)
pTOI07 (ORF, - )
pTOI08 (BPST.1)

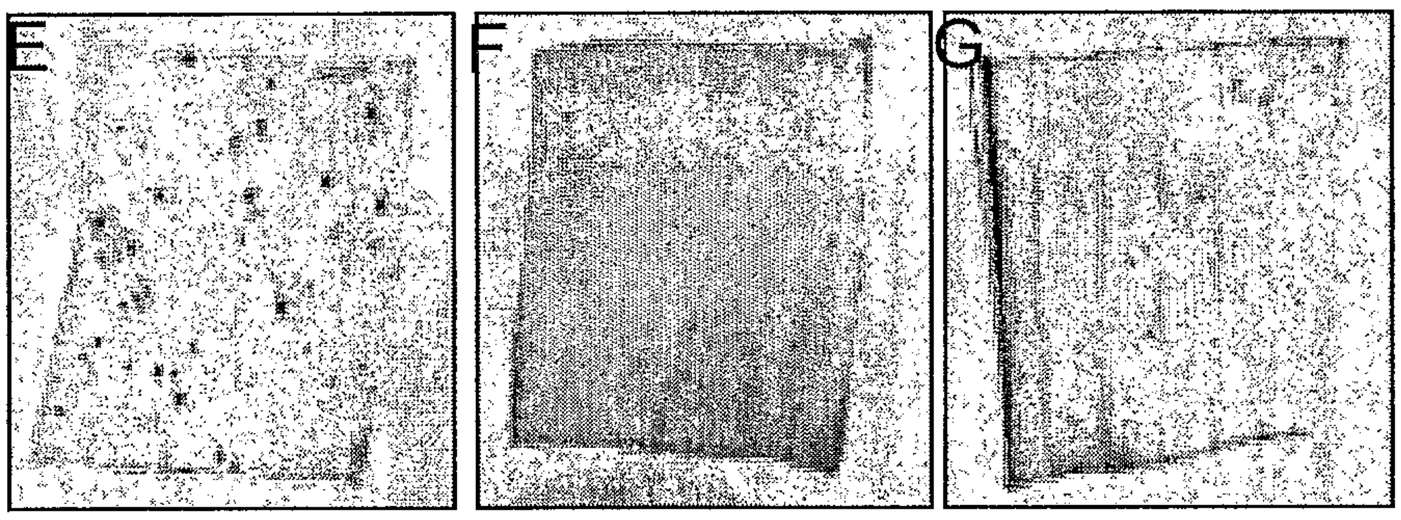

E
F
G
pTOI09 (BPST.2)
pTOI10 (BPST.3)
pTOI11 (BPST.4)

Fig. 14

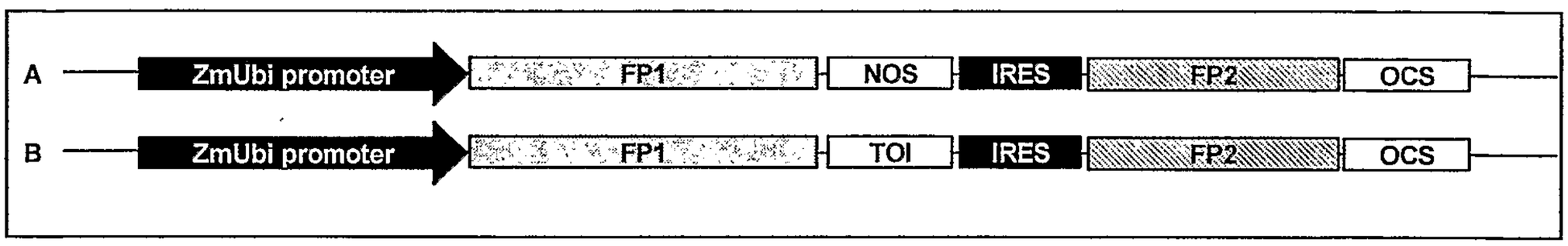

A
ZmUbi promoter
FP1
NOS
IRES
FP2
OCS
B
ZmUbi promoter
FP1
TOI
IRES
FP2
OCS

Fig. 15

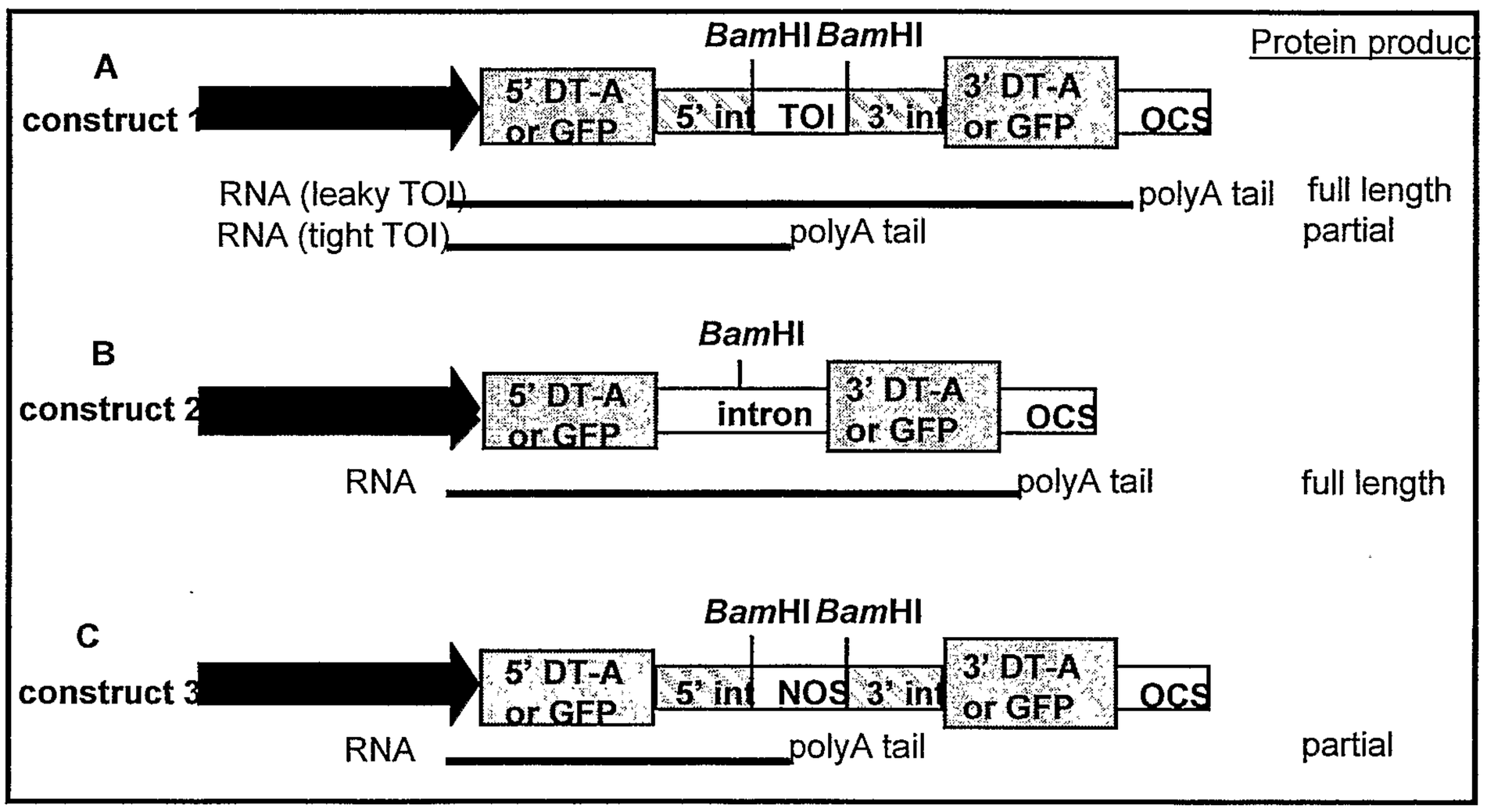

A
construct 1
BamHI BamHI
Protein product
5' DT-A or GFP
5' int
TOI
3' int
3' DT-A or GFP
OCS
RNA (leaky TOI)
polyA tail
full length
RNA (tight TOI)
polyA tail
partial
B
construct 2
BamHI
5' DT-A or GFP
intron
3' DT-A or GFP
OCS
RNA
polyA tail
full length
C
construct 3
BamHI BamHI
5' DT-A or GFP
5' int
NOS
3' int
3' DT-A or GFP
OCS
RNA
polyA tail
partial

Fig. 16

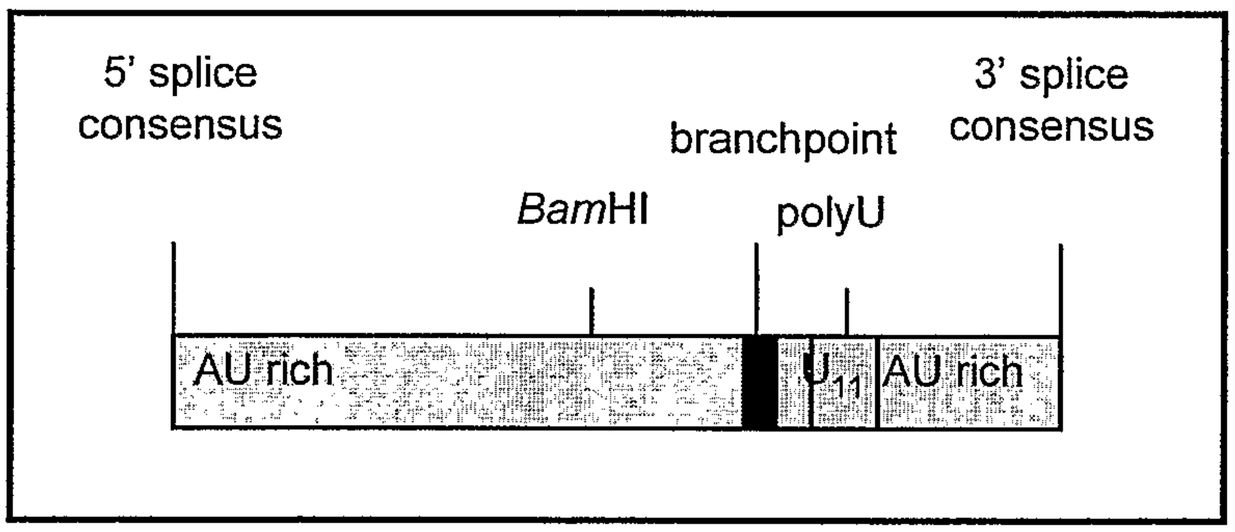

5' splice consensus
branchpoint
3' splice consensus
BamHI
polyU
AU rich
U11
AU rich

Fig. 17

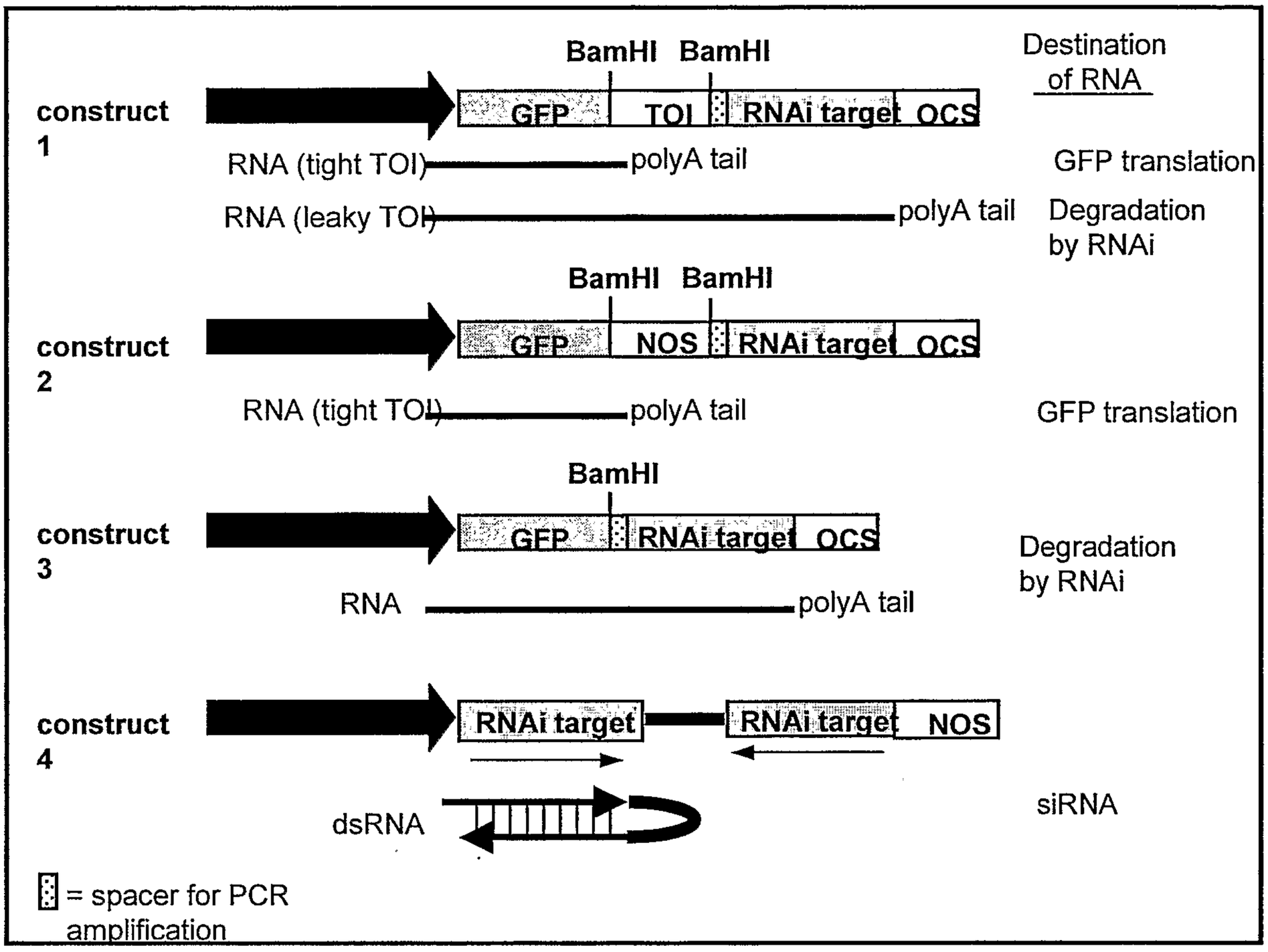
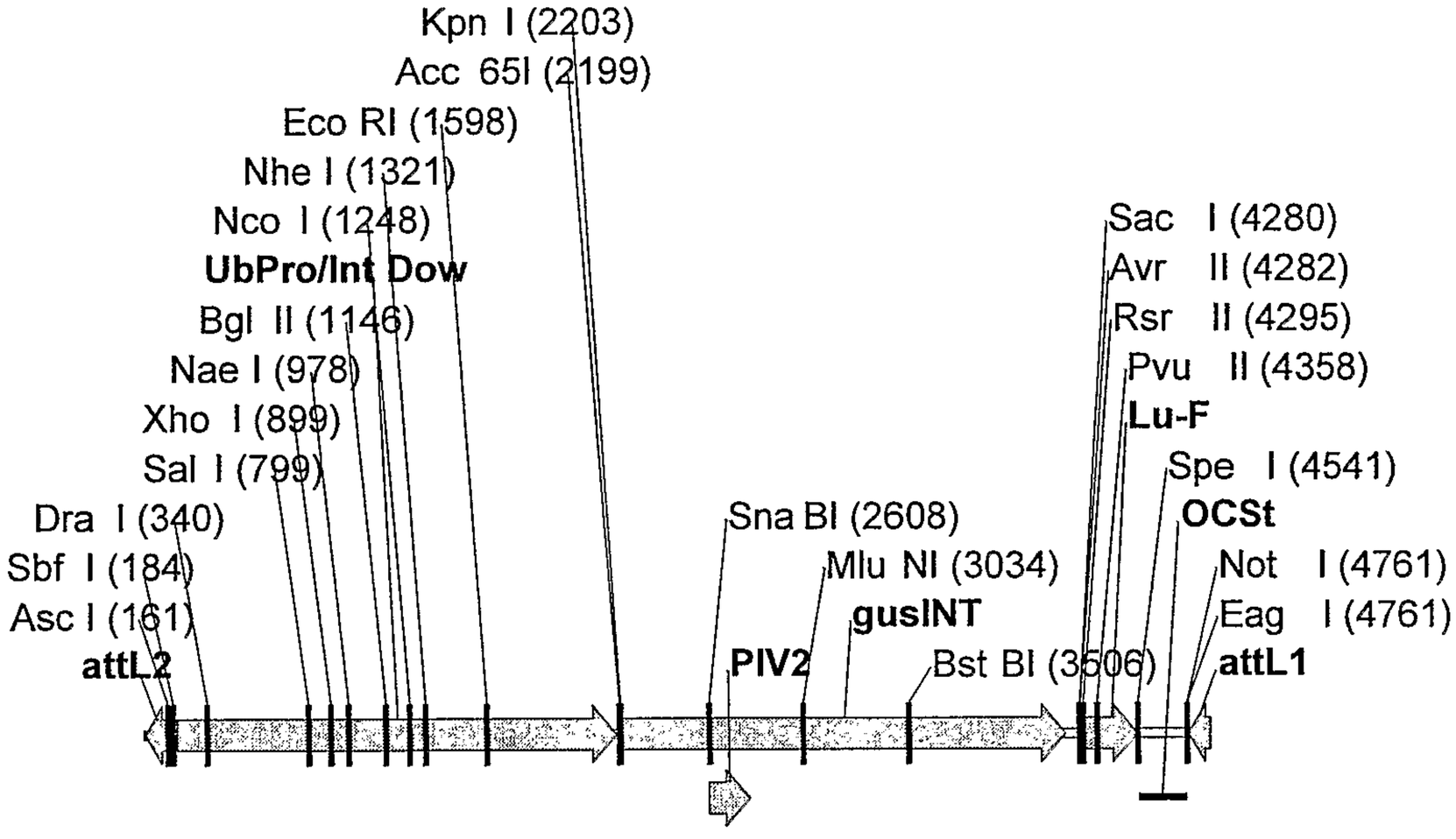
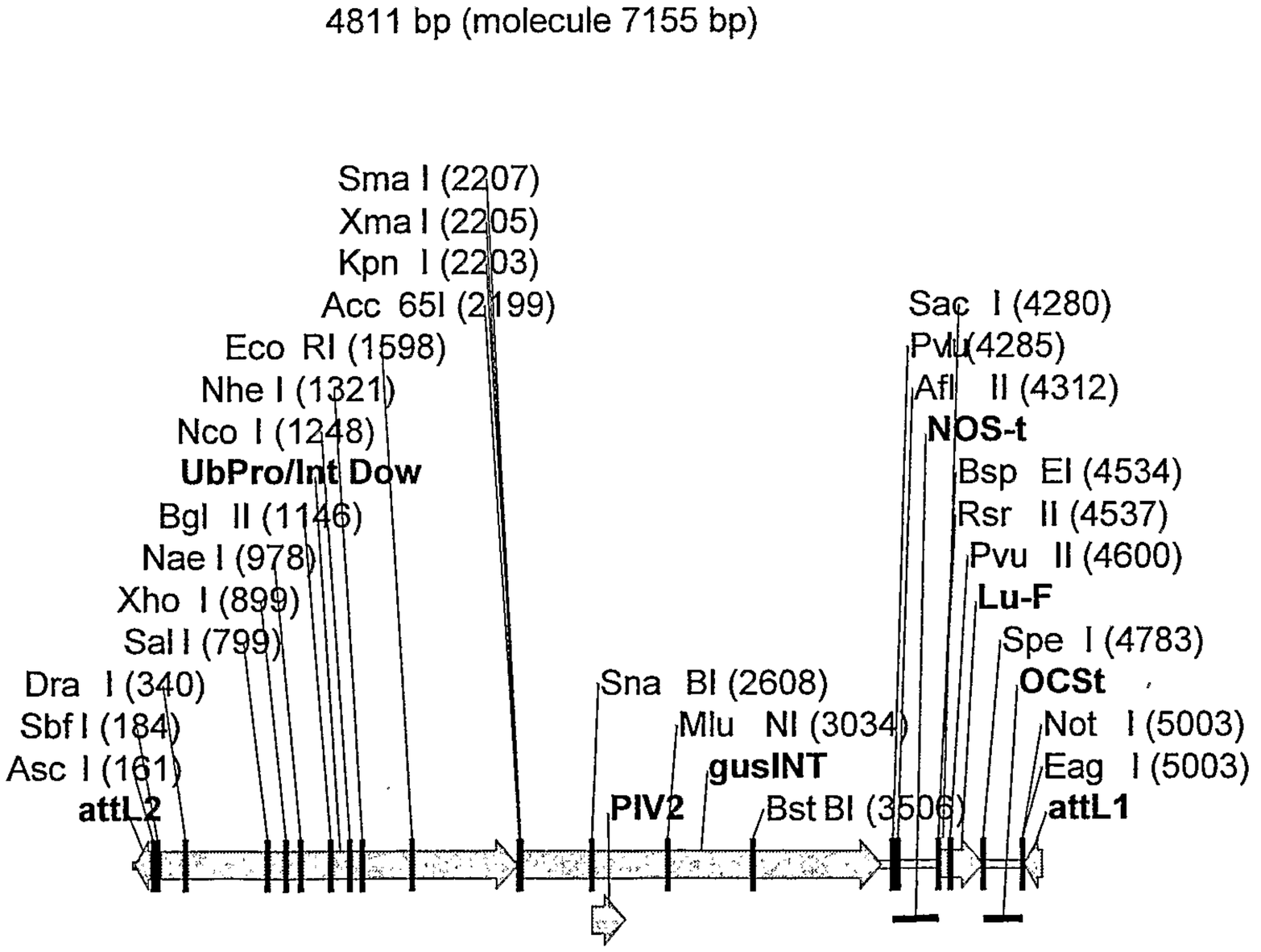

BamHI BamHI
Destination of RNA
construct 1
GFP
TOI
RNAi target
OCS
RNA (tight TOI)
polyA tail
GFP translation
RNA (leaky TOI)
polyA tail
Degradation by RNAi
BamHI BamHI
construct 2
GFP
NOS
RNAi target
OCS
RNA (tight TOI)
polyA tail
GFP translation
BamHI
construct 3
GFP
RNAi target
OCS
RNA
polyA tail
Degradation by RNAi
construct 4
RNAi target
RNAi target
NOS
dsRNA
siRNA
= spacer for PCR amplification

Fig. 18




Fragment of pTOI 03
4811 bp (molecule 7155 bp)




Fragment of pTOI 04
5053 bp (molecule 7397 bp)

Fig. 19

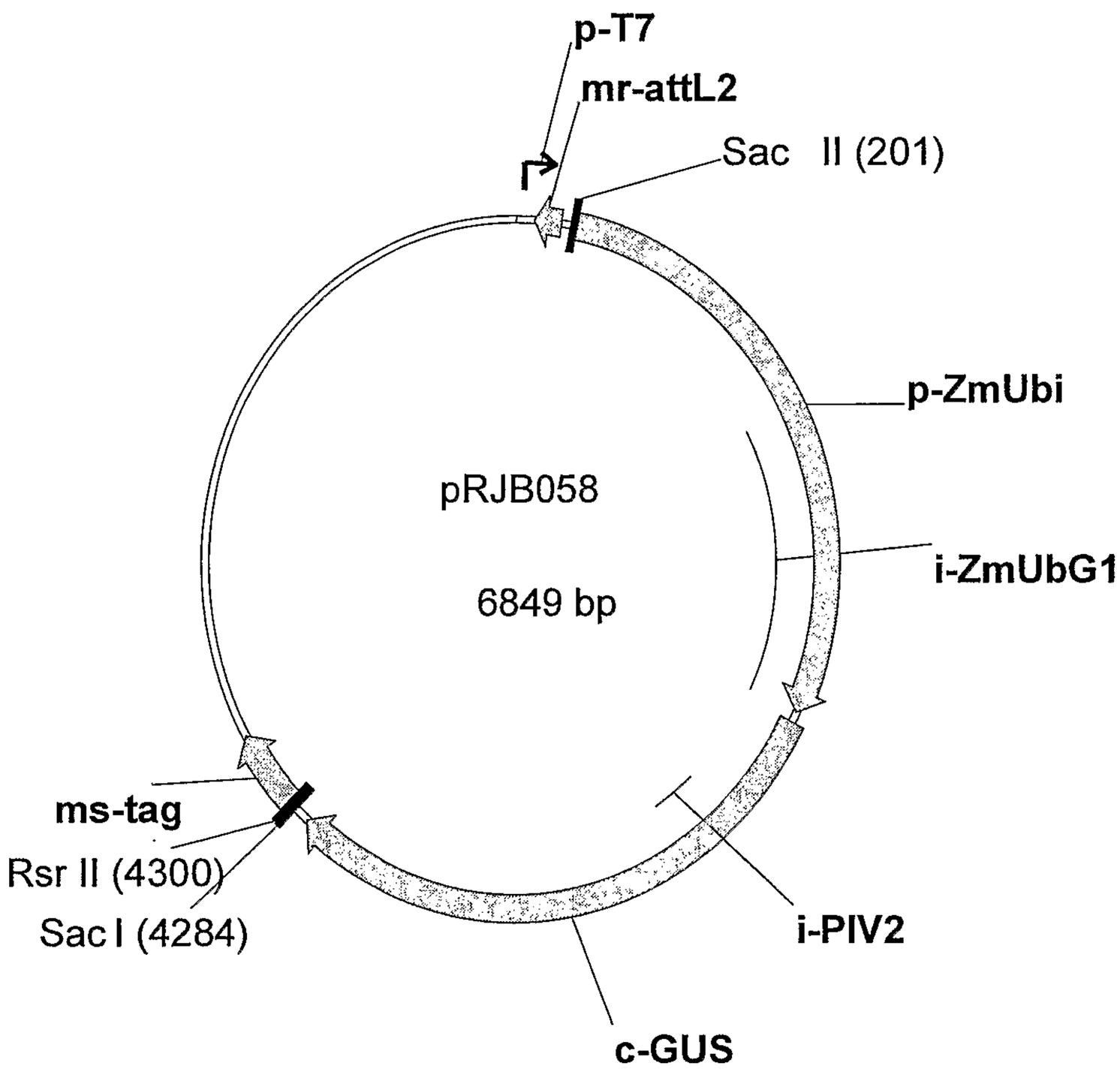
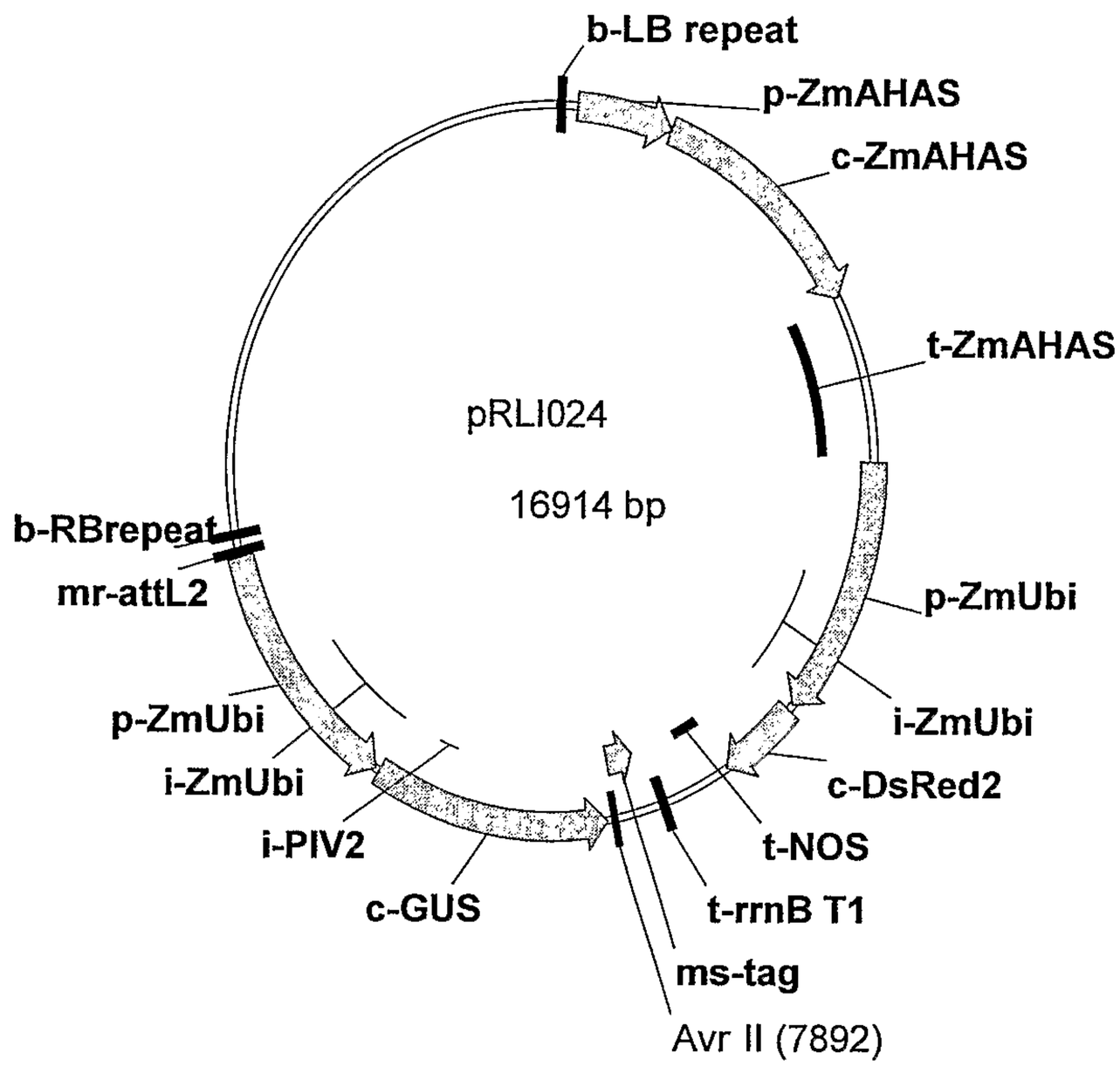

p-T7
mr-attL2
Sac II (201)
p-ZmUbi
pRJB058
6849 bp
i-ZmUbG1
ms-tag
Rsr II (4300)
Sac I (4284)
i-PIV2
c-GUS

Fig. 20

A


b-LB repeat
p-ZmAHAS
c-ZmAHAS
t-ZmAHAS
pRLI024
16914 bp
b-RBrepeat
mr-attL2
p-ZmUbi
i-ZmUbi
p-ZmUbi
i-ZmUbi
i-PIV2
c-GUS
t-NOS
t-rrnB T1
ms-tag
Avr II (7892)
c-DsRed2

B

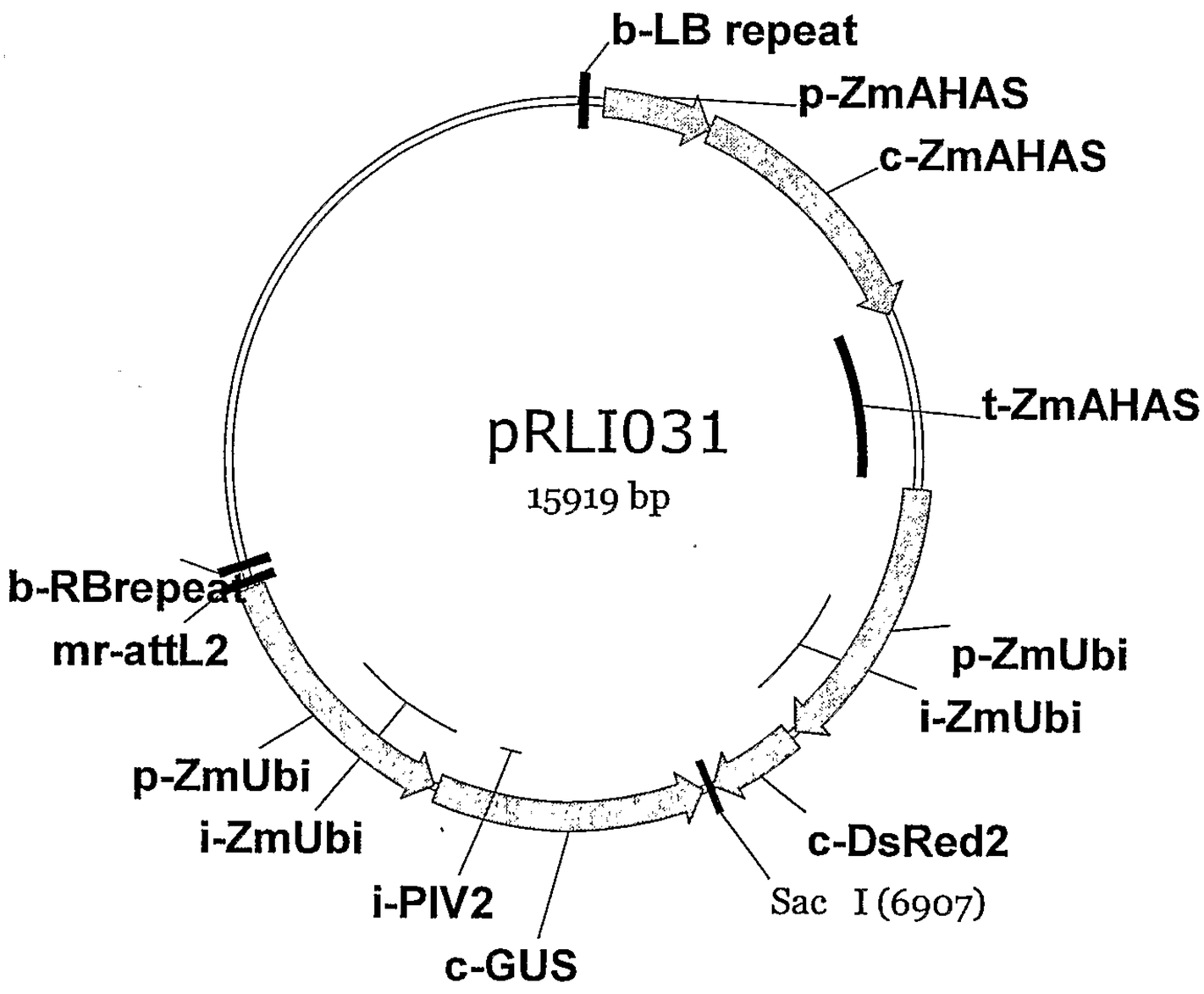
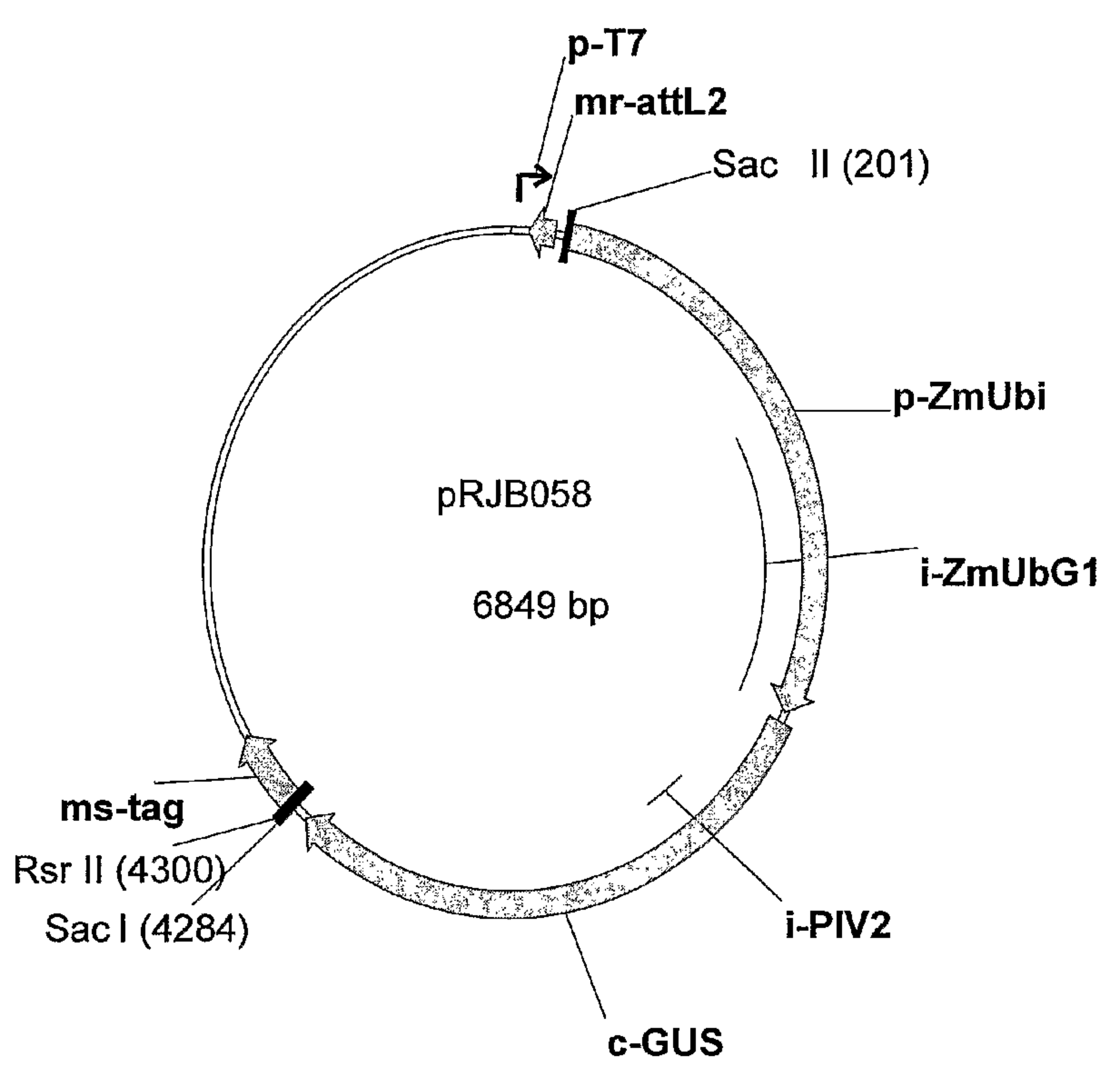

b-LB repeat
p-ZmAHAS
c-ZmAHAS
t-ZmAHAS
pRLI031
15919 bp
b-RBrepeat
mr-attL2
p-ZmUbi
i-ZmUbi
p-ZmUbi
i-ZmUbi
i-PIV2
c-GUS
Sac I (6907)
c-DsRed2

Fig. 21


p-T7
mr-attL2
Sac II (201)
p-ZmUbi
pRJB058
6849 bp
i-ZmUbG1
ms-tag
Rsr II (4300)
Sac I (4284)
i-PIV2
c-GUS