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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 January 2011 (27.01.2011)

PCT

(10) International Publication Number
WO 2011/009700 A1

(51) International Patent Classification:
C12N 15/81 (2006.01) *C12N 15/56* (2006.01)
C12N 15/00 (2006.01) *C12N 9/60* (2006.01)
C12N 1/16 (2006.01)

(21) International Application Number:
PCT/EP2010/059390

(22) International Filing Date:
1 July 2010 (01.07.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09166074.6 22 July 2009 (22.07.2009) EP

(71) Applicant *(for all designated States except US)*: **DSM IP ASSETS B.V.** [NL/NL]; Het Overloon 1, NL-6411 TE Heerlen (NL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PEIJ, VAN, Noël Nicolaas Maria Elisabeth** [NL/NL]; Poldermeesterstraat 7, NL-2645 KJ Delfgauw (NL). **PEL, Herman Jan** [NL/NL]; Rijksweg 131, NL-3784 LV Terschuur (NL). **WENZEL, Thibaut, José** [NL/NL]; Constantijn Huygenslaan 45, NL-2332 TS Leiden (NL). **RIEMENS, Adriana Marina** [NL/NL]; Leharlaan 9, NL-2625 ZE Delft (NL). **LANGE, DE, Ilse** [NL/NL]; Kersengaarde 203, NL-3225 GR Hellevoetsluis (NL).

(74) Agents: **HAAFT, TEN, Petrus Johannes Fredrik** et al.; DSM Intellectual Property, Delft Office (600-0240), P.O. Box 1, NL-2600 MA Delft (NL).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,

[Continued on next page]

(54) Title: IMPROVED HOST CELL FOR THE PRODUCTION OF A COMPOUND OF INTEREST

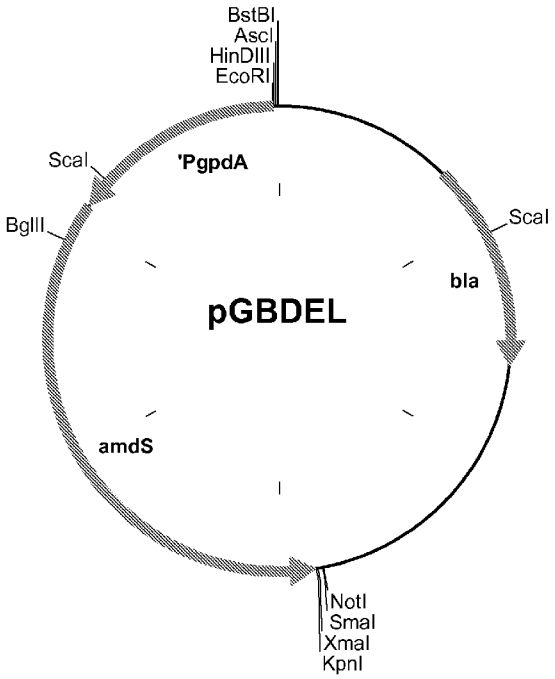


Figure 1

(57) Abstract: The present invention relates to a recombinant host cell for the production of a compound of interest. The invention further relates to a method for the production of such host cell. The invention further relates to the production of a compound of interest. The invention further relates to isolated polynucleotides and vectors and host cells comprising said polynucleotides.

WO 2011/009700 A1

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

— *with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description (Rules 13bis.4(d)(i) and 48.2(a)(viii))*

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

IMPROVED HOST CELL FOR THE PRODUCTION OF A COMPOUND OF INTEREST

Field of the invention

The present invention relates to a recombinant host cell for the production of a compound of interest. The invention further relates to a method for the production of such host cell. The invention further relates to the production of a compound of interest. The invention further relates to isolated polynucleotides and vectors and host cells comprising said polynucleotides.

Background of the invention

The present invention relates to a recombinant host cell for the production of a compound of interest.

Such host cells are inter alia known from WO1998/046774 and WO1998/46772, wherein the host cell comprises a polynucleotide of interest in one of at least two substantially homologous DNA domains of the chromosome(s) of said host cell and wherein the copy number of the polynucleotide of interest is increased by means of gene conversion.

However, it has been demonstrated that obtaining a high-copy gene conversion strain, such as those obtained by the method according to WO1998/046774 and WO1998/46772, is often considered to be laborious. It would be therefore be advantageous if the method could be improved in order to enable less laborious construction of a high-copy gene conversion strain.

Description of the Figures

Figure 1 depicts a physical map of replacement vector pGBDEL. Indicated are the multiple cloning sites for cloning flanking regions relative to the amdS marker.

Figure 2 depicts a physical map of replacement vector pDEL-HDFA for use to inactivate the *hdfA* gene in *Aspergillus niger (A. niger)*. The replacement vector comprises the *hdf*A flanking regions, the *amd*S marker and *E. coli* DNA. The *E. coli*

DNA can be removed by digestion with restriction enzymes AscI and NotI, prior to transformation of the *A. niger* strains.

Figure 3 depicts the strategy used to delete a gene or gene fragment from a host strain. The DNA construct used comprises the *amd*S selection marker flanked by homologous regions (5' and 3') of the gene to be deleted (1). This construct integrates through double homologous recombination (**X**) at the corresponding genomic locus (2) and replaces the genomic gene copy (3). Subsequently, recombination over the direct repeats (**U**) removes the *amd*S marker, resulting in precise excision of the gene or gene fragment to be deleted (4).

Figure 4 depicts physical maps of the *gla*A (glucoamylase) loci in parental *A. niger* strain CBS 124.903 and the three truncated "X-marked" Δ*gla*A loci in recombinant strain GBA 300 ("X" stands for a *Bam*HI, *Sal*I or *Bgl*II restriction site).

Figure 5 depicts a schematic view of the three Δ*gla*A loci in the *gla*A DNA amplicons of *A. niger* CBS 124.903 derived recombinant strain GBA 300, each locus marked by a different restriction site (*Bam*HI, *Sal*I or *Bgl*II). The *gla*A loci each differ in length by approximately 20 to 60 bp, to be able to visualize the different truncated *gla*A loci by the PCR-based DNA-flag test.

Figure 6 depicts a physical map of vector pGBGLA-65 for use to adapt a substantially homologous DNA domain. Basically, this vector contains an *amd*S selection marker in between two P_{glaA} fragments and the 3'*gla*A and 3"*gla*A region. The 3'*gla*A and 3"*gla*A region are used for targeting and integration into the corresponding genomic region of the Δ*gla*A loci. The two P_{glaA} fragments are used for excision of the amdS selection marker upon counterselection as in Figure 3. One PglaA fragment is a truncated P_{glaA} promoter fragment (missing the last 600 bp of the P_{glaA} promoter 3' of the *Mlu*I site), which remains present in the genome after amdS counterselection.

Figure 7 depicts physical maps of the three Δ*gla*A loci in the *gla*A DNA amplicons in a strain with an adapted *Bam*HI amplicon. The *Bam*HI amplicon was

adapted by introduction of a truncated homologous glucoamylase promoter fragment in between the homologous 3' *gla*A and 3" *gla*A loci of the genomic *gla*A locus.

Figure 8 depicts a physical map of a novel acetamidase selection marker and adapted *Bam*HI amplicon targeting vector pGBAAS-3. The *gla*A promoter functions as targeting region and as promoter for the *amd*S gene

Figure 9 depicts a physical map of a novel acetamidase selection marker and adapted *Bam*HI amplicon targeting vector pGBAAS-4. The (truncated) *gla*A promoter functions as targeting region and the gpdA promoter drives the expression of the *amd*S gene

Figure 10 depicts a physical map of a novel expression and adapted *Bam*HI amplicon targeting vector pGBTOP-11. The *Hin*DIII restriction enzyme can be used to linearize the vector and remove the *E. coli* part

Figure 11 depicts a physical map of a novel expression and adapted *Bam*HI amplicon targeting vector pGBTOP-12. The *Not*I restriction enzyme can be used to linearize the vector and remove the *E. coli* part

Figure 12 depicts a physical map of replacement vector pDEL-AMYBII. Indicated are the 5' *amy*BII flanking region, the 3' *amy*BII flanking regions relative to the *amd*S marker. The sequence of the 3'- sequences of *amy*BII overlap at least a few hundred bp. The *E. coli* DNA can be removed by digestion with restriction enzyme *Not*I, prior to transformation of the host strain.

Figure 13 depicts a physical map of replacement vector pGBDEL-FUM3. Indicated are the 5' *fum*B region with part of the *fum*B coding sequence and the 3' *fum*B flanking regions relative to the *amd*S marker. The sequence of the 5'- sequences of *fum*B overlap at least a few hundred bp. The *E. coli* DNA can be removed by digestion with restriction enzyme *Asc*I and *Not*I, prior to transformation of the host strain.

Figure 14 depicts a physical map of replacement vector pGBDEL-OCH2. Indicated are part of the An15g07860 gene sequence and part of the An15g07930 gene sequence as flanking regions relative to the *amd*S marker. The sequence of the An15g07930 gene overlap at least a few hundred bp to allow counter selection by homologous recombination. The *E. coli* DNA can be removed by digestion with restriction enzyme *Asc*I and *Not*I, prior to transformation of the host strain.

Figure 15 depicts a physical map of replacement vector pGBDEL-PRT2. Indicated are the 5' prtT flanking region, the 3' prtT flanking regions relative to the amdS marker. The sequence of the prtT 3' sequences overlap at least a few hundred bp. The *E. coli* DNA can removed by digestion with restriction enzyme BstBI and XmaI, prior to transformation of the host strain.

Figure 16 depicts a physical map of replacement vector pGBDEL-Sec61*. Indicated are a Sec61* mutant gene and a 3' Sec61* mutant gene fragment relative to the *amd*S marker. The *E. coli* DNA can be removed by digestion with restriction enzyme *Asc*I and *Not*I, prior to transformation to the host strain.

Figure 17 depicts a physical map of the PLA2 expression vector pGBTOPPLA-2a. Indicated are the *gla*A flanking regions relative to the *gla*A promoter, the truncated *gla*A gene and the pro-*pla*2 coding sequence. The *E. coli* DNA can be removed by digestion with restriction enzyme *Not*I, prior to transformation to the host strain.

Figure 18 depicts a physical map of the PLA2 expression vector pGBTOPPLA-2b. Indicated are the *gla*A promoter, the truncated *gla*A gene, the pro-*pla*2 coding sequence and the 3'*gla*A flanking sequence. The *E. coli* DNA can be removed by digestion with restriction enzyme *Not*I, prior to transformation to the host strain.

Figure 19 depicts a physical map of expression vector pGBFINFUA-2. Indicated are the *gla*A flanking regions relative to the *gla*A promoter and the *A. niger amy*B codon optimized sequence. The *E. coli* DNA can be removed by digestion with restriction enzyme *Not*I, prior to transformation of the host strain.

Figure 20 depicts a physical map of the FUA expression vector pGBTOPFUA-3. Indicated are the *gla*A promoter, the truncated *gla*A gene, the *A. niger amy*B codon optimized sequence and the 3'*gla*A flanking sequence. The *E. coli* DNA can be removed by digestion with restriction enzyme *Not*I, prior to transformation to the host strain.

Figure 21 depicts a schematic representation of integration through single homologous recombination in a strain with an adapted *Bam*HI amplicon (as for example GBA 303-based strains in the examples). The expression vector comprises a selectable amdS marker with specific targeting sequences and a gene of interest. The targeting sequences allow specific genomic targeting to the adapted amplicon with the highest frequency of gene conversion ("Smart integration strategy").

Detailed description of the invention

Analysis revealed surprisingly that where a host cell, such as those described in WO1998/046774 and WO1998/46772, comprises at least two substantially homologous DNA domains, the gene conversion frequencies of the DNA domains are not necessarily equal. Some of these DNA domains revealed to have a gene conversion frequency that is substantially lower than that of other substantially homologous DNA domains. In such case, it is a disadvantage if the polynucleotide of interest is integrated in a DNA domain that has a low gene conversion frequency and not in a DNA domain that has a high gene conversion frequency. The DNA domain with the higher gene conversion frequency will to a large extent out-compete the DNA domain with the lower gene conversion frequency (which comprises the polynucleotide of interest). Consequently, it will be more laborious to obtain a high-copy gene conversion host cell from a cell that has the polynucleotide of interest integrated in a DNA domain with lower gene conversion frequency.

It would be therefore be advantageous if the polynucleotide of interest can be comprised in a substantially homologous DNA domain that has a substantially higher gene conversion frequency since this would enable less laborious construction of a host strain comprising multiple copies of the polynucleotide of interest (high-copy gene conversion strain).

Consequently, there is a need for a host cell wherein the polynucleotide of interest preferentially integrates in a homologous DNA domain that has a substantially higher gene conversion frequency. It is therefore an object of the present invention to provide such host cell.

Accordingly, the present invention provides for a recombinant host cell for the production of a compound of interest, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest wherein at least one of the at least two substantially homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from. Said recombinant host cell is herein referred to as the recombinant host cell according to the invention.

The present invention also provides for a method for the production of a recombinant host cell for the production of a compound of interest, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest, comprising adapting a substantially homologous DNA domain with higher gene conversion frequency compared to another substantially homologous DNA domain to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from.

The present invention also provides for a method for the production of a compound of interest, comprising:

a. cultivating a host cell according to the invention under conditions conducive to the production of said compound; and

b. recovering the compound of interest from the cultivation medium.

The present invention also provides for a method for the production of a compound of interest comprising

a. cultivating a recombinant host cell under conditions conducive to the production of said compound, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest, wherein at least one of the at least two substantially

homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, and wherein at least two of the substantially homologous DNA domains have at least one copy of a polynucleotide of interest integrated; and

b. recovering the compound of interest from the cultivation medium.

Accordingly, a recombinant host cell according to the invention comprises at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest wherein at least one of the at least two substantially homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from. Said substantially homologous DNA domains suitable for integration of one or more copies of said polynucleotide of interest preferably are substantial duplicates of each other; more preferably said DNA domains are amplicons.

In the context of the present invention, the term “recombinant” refers to any genetic modification not exclusively involving naturally occurring processes and/or genetic modifications induced by subjecting the host cell to random mutagenesis but also gene disruptions and/or deletions and/or specific mutagenesis, for example. Consequently, combinations of recombinant and naturally occurring processes and/or genetic modifications induced by subjecting the host cell to random mutagenesis are construed as being recombinant.

The host cells according to the invention comprise in their genome at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest. In wild-type strains these DNA domains may occur in a single copy or in multiple copies in the genome of said cell. Examples of multiple copy DNA domains occurring naturally are rDNA domains. Accordingly, rDNA domains may preferably be used or not be used in the present invention. Strains containing multiple copies of these substantially homologous DNA domains, as provided by the present invention, can e.g. be obtained from strains comprising a single copy of a DNA domain by classical strain improvement by selecting for strains with improved production of the gene products encoded by the gene(s) in said DNA domains. Frequently, such production improvements are the result of amplification of a DNA domain in the

selected strains. Such amplified DNA domains are also referred to as amplicons hereinafter. Examples of host cells comprising such amplicons are e.g. described in van Dijck et al, 2003, Regulatory Toxicology and Pharmacology 28; 27-35: *On the safety of a new generation of DSM Aspergillus niger enzyme production strains*. In van Dijck et al, an *Aspergillus niger* strain is described that comprises 7 amplified glucoamylase gene loci, i.e. amplicons. Although the present invention preferably uses such amplicons as DNA domains for the integration of the recombinant DNA molecules, the invention is by no means limited thereto. In fact, any DNA domain of which two or more substantially homologous versions occur in the genome of a host cell can be used as long as two functional criteria are fulfilled: 1) the DNA domain should be suitable for accepting the integration of a polynucleotide of interest; 2) the DNA domain should be capable of recombination with the other substantially homologous versions of the DNA domain in the fungal genome in order to achieve multiplication of the integrated recombinant DNA molecule through gene conversion.

In order to meet the first criterion, a DNA domain must be of sufficient length in order to allow targeting of the polynucleotide of interest into the DNA domain through homologous recombination. For this purpose a DNA domain should comprise at least 100 bp, preferably at least 500 bp, more preferably at least 1 kb and more preferably at least 2 kb. Preferably, the suitability of a DNA domain for integration therein of a polynucleotide of interest is furthermore determined by the requirement that integration into the DNA domain should not disrupt functions that are essential for the viability of the host cell in question.

The second functional criterion, i.e. the capability of recombination with the other substantially homologous versions of the DNA domain in the genome of the host cell, is required for allowing gene conversions between the different versions of the DNA domain. The minimal requirement for this purpose is that each version of the DNA domain is flanked on either end of the DNA domain by DNA sequences that are sufficiently homologous to the corresponding flanking sequences of the other version of the DNA domains so as to allow homologous recombination between the flanking sequences. The result of this homologous recombination being a gene conversion wherein one of the versions of the DNA domain is replaced by a duplicated version of the other DNA domain containing the integrated recombinant DNA molecule. The minimum requirements regarding length and extend of homology of the flanking

sequences still allowing gene conversion may vary depending on the organism in question. Probably, a minimum length of 100 bp with an overall homology at least 60% will still allow gene conversion. Obviously, the frequency of gene conversion will increase with increasing length and homology of the flanking sequences. Preferably the different DNA domains comprise flanking sequences of at least 1 kb which share at least 80% homology. Even more preferably the different DNA domains comprise flanking sequences of at least 5 kb which share at least 95% homology. In a more preferable embodiment the different DNA domains comprise flanking sequences of at least 10 kb which share at least 99% homology. In a most preferable embodiment, the different DNA domains comprise flanking sequences of at least 50 kb which share at least 99% homology. Examples of substantially homologous DNA domains which are not perfect copies of each other, designated herein substantial duplicates, are allelic variants, gene families and/or genes encoding iso-enzymes. Most preferred are substantially homologous DNA domains which are exact copies of each other, differing at most in the presence of the integrated polynucleotide of interest. Examples of such identical DNA domains are amplicons. Alternatively, artificial DNA domains may be created using DNA synthesis and/or recombinant DNA technology, and such artificial DNA domains may then be introduced and amplified in the desired fungus. Accordingly, such artificial DNA domains may preferably be used or not be used in the present invention.

The overall length of the substantially homologous DNA domains, i.e. including the flanking sequences may vary from less than 1 kb to several hundreds of kb's, e.g. previously it has been demonstrated that the length of the DNA domains ranges from few kb's per unit for the three TAKA amylase genes present in the *Aspergillus oryzae* parental strain IF04177 (disclosed in WO9812300 and U.S. Pat. No. 5,766,912 as A1560), which have been inactivated in *A. oryzae* BECh2 (disclosed in WO0039322) until about 57 kb per unit for the amplified penicillin cluster in *Penicillium chrysogenum* and more than 80 kb per unit for the amplified glaA locus of *Aspergillus niger* (see WO9846772).

Gene conversion can be monitored by identifying host cells in which the DNA domain comprising the polynucleotide of interest is multiplied. Any method known to the person skilled in the art can be used, e.g. simply screening for host cells with higher production levels of the product encoded by polynucleotide of interest. Also quantitative

methods for determination of DNA copy numbers such as for example analysis of the genotype of host cells by e.g. the "DNA-flag" test as outlined here below, quantitative PCR, Southern analysis or comparative genome hybridization or for determination of RNA levels, such as for example RT-PCR, Northern analysis or GeneChip or microarray analysis can be used.

According to the invention, each version of a substantially homologous DNA domain is preferably distinguished from the other versions of the DNA domains in the filamentous fungus by means of a unique sequence tag for identification purposes. Such sequence tags allow detection and monitoring gene conversions between the different DNA domains, thus facilitating the screening and/or selection of gene convertants with a desired genotype. Any form of sequence tags can be used as long as they allow detection of the different versions of the DNA domain: e.g. ranging from restriction sites that are detected on Southern blots to complete selectable marker genes providing an easy assayable phenotype. A particularly useful sequence tag method is described in WO9846772, WO9846774 and in van Dijck et al, *supra*. This method allows detecting each of the DNA domains in a single PCR using a single pair of oligonucleotide PCR primers. The DNA domains are modified in such a way that in the PCR each version of the DNA domains produces a PCR fragment with a unique length. The length and intensity of the obtained PCR fragments are indicative for the presence and copy number of each of the DNA domains, respectively. This form of sequence tag, referred to as "DNA flags", allows for rapid analysis of the genotype of large numbers of convertant colonies, in order to obtain a convertant with the desired genotype.

According to the invention, one of the at least two substantially homologous DNA domains of the host cell according to the invention is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from. Preferably, a targeting DNA domain of one of the at least two substantially homologous DNA domains of the host cell is adapted to allow a more specific targeting of the recombinant DNA molecule / polynucleotide of interest into the adapted targeting DNA domain through homologous recombination. If a targeting DNA sequence or DNA domain occurs more than once in a host cell, homologous recombination of a polynucleotide of interest is found with multiple targeting DNA domains of a host cell. It is possible to adapt a targeting DNA domain in

one of the substantially homologous DNA domains to support more specific targeting to that adapted targeting DNA domain without affecting the recombination with other substantially homologous versions of the DNA domain in the host. For optimal targeting preference, adaptation of the DNA sequences used for targeting is done in the recombinant DNA molecule / polynucleotide of interest and in the host. As such, a matching set of host and integration vector are obtained.

Preferably, adapting the targeting DNA domain is executed by introduction of a sequence with enhanced integration preference into the substantially homologous DNA domain. Preferably, said sequence with enhanced integration preference is introduced by gene replacement. Accordingly, the present invention also provides for a recombinant host cell for the production of a compound of interest, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest wherein at least one of the at least two substantially homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, wherein the adapted substantially homologous DNA domain comprises a targeting DNA domain, wherein said targeting DNA domain comprises a sequence with enhanced integration preference. More preferably, adapting the targeting domain is executed according to the examples herein.

According to the invention, a targeting sequence of the adapted DNA domain used for targeting is preferably adapted to be unique within the genome of the host. Preferably, the targeting sequence adapted to be unique within the genome of the host comprises a sequence selected from the group of: heterologous DNA polynucleotide sequences, DNA polynucleotide sequences previously removed from the host and re-introduced, homologous DNA polynucleotide fragments assembled in a different order compared to the wild-type situation in a host, and artificial DNA polynucleotides.

According to the invention, one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain are preferably unique within the genome of the host. Preferably, one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain comprises a sequence selected from the group of: heterologous DNA polynucleotide sequences, DNA polynucleotide sequences previously removed from the host and re-introduced,

homologous DNA polynucleotide fragments assembled in a different order compared to the wild-type situation in a host and artificial DNA polynucleotides.

Preferably, both a targeting sequence of the adapted DNA domain is adapted to be unique within the genome of the host and one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain are unique within the genome of the host.

According to the invention, a sequence used for targeting to the adapted DNA domain preferably is a genetic element used for introduction and / or expression of a polynucleotide of interest. Examples of genetic elements are a gene, a promoter, a terminator, a cDNA, an intron, an intergenic area, or a part and / or a combination thereof. More preferably, a sequence used for targeting to the adapted DNA domain comprises one or more heterologous genetic elements for introduction and / or expression of a polynucleotide of interest. Even more preferably, one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain comprises one or more heterologous genetic elements for introduction and / or expression of a polynucleotide of interest. Even more preferably, a sequence used for targeting to the adapted DNA domain comprises one or more *A. niger* genetic elements for introduction and / or expression of a polynucleotide of interest. Even more preferably, one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain comprises one or more *A. niger* genetic elements for introduction and / or expression of a polynucleotide of interest. Even more preferably, a sequence used for targeting to the adapted DNA domain comprises *A. niger gla*A glucoamylase genetic elements and / or flanking regions for introduction and / or expression of a polynucleotide of interest. even more preferably, one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain comprises *A. niger gla*A glucoamylase genetic elements and / or flanking regions for introduction and / or expression of a polynucleotide of interest. Even more preferably, a sequence used for targeting to the adapted DNA domain comprises the *A. niger gla*A glucoamylase promoter and / or glucoamylase terminator regions for introduction and / or expression of a polynucleotide of interest. Even more preferably, one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain comprises the *A. niger gla*A glucoamylase promoter and / or

glucoamylase terminator regions for introduction and / or expression of a polynucleotide of interest. Even more preferably, a sequence used for targeting to the adapted DNA domain comprises the *A. niger gla*A glucoamylase terminator region and *gla*A glucoamylase promoter in a reverse order compared to the wild-type situation for introduction and / or expression of a polynucleotide of interest. Even more preferably, one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain comprises the *A. niger gla*A glucoamylase terminator region and *gla*A glucoamylase promoter in a reverse order compared to the wild-type situation for introduction and / or expression of a polynucleotide of interest. Even more preferably, a sequence used for targeting to the adapted DNA domain comprises the *A. niger gla*A glucoamylase terminator region and *gla*A glucoamylase promoter as identified in SEQ ID NO: 2 for introduction and expression of a polynucleotide of interest. Even more preferably, one or both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain comprises the *A. niger gla*A glucoamylase terminator region and *gla*A glucoamylase promoter as identified in SEQ ID NO: 2 for introduction and expression of a polynucleotide of interest. Most preferably, the sequences used for targeting to the adapted DNA domain and both flanking sequences of the polynucleotide of interest used for targeting to the adapted DNA domain are the same and comprise the *A. niger gla*A glucoamylase terminator region and *gla*A glucoamylase promoter as identified in SEQ ID NO: 2 for introduction and expression of a polynucleotide of interest.

Accordingly, the present invention provides for a polynucleotide comprising a polynucleotide according to SEQ ID NO: 2 or a functional fragment thereof and vectors and host cells comprising a polynucleotide comprising a polynucleotide according to SEQ ID NO: 2 or a functional fragment thereof. Functional is here to be construed as having promoter activity and a being a targeting sequence with enhanced integration preference. Preferably, a functional fragment of a polynucleotide according to SEQ ID NO: 2 comprises at least 100 bp, more preferably at least 200bp, more preferably at least 300bp, more preferably at least 500bp, more preferably at least 600bp, more preferably at least 700bp, more preferably at least 800bp, more preferably at least 900bp, more preferably at least 1000bp, more preferably at least 1200bp, more preferably at least 1500bp, more preferably at least 2000bp, more preferably at least

3000bp, and even more preferably at least 4000bp. Preferably, the functional fragment has a match percentage, i.e. positional identity of at least about 50%, more preferably of at least about 60%, even more preferably of at least about 70%, even more preferably of at least about 80%, even more preferably of at least about 85%, even more preferably of at least about 90%, even more preferably of at least about 95%, even more preferably of at least about 97%, even more preferably of at least about 98%, even more preferably of at least about 99% identity, and even preferably, the match percentage i.e. identity is equal to 100%.

Most preferably, the polynucleotide comprising a polynucleotide according to SEQ ID NO: 2 comprises SEQ ID NO: 3 or a functional fragment thereof. Functional is here to be construed as having promoter activity and a being a targeting sequence with enhanced integration preference.

Accordingly, the present invention provides for a polynucleotide comprising a polynucleotide according to SEQ ID NO: 3 or a functional fragments thereof and vectors and host cells comprising a polynucleotide comprising a polynucleotide according to SEQ ID NO: 3 or functional fragments thereof. Preferably, a functional fragment of SEQ ID NO: 3 comprises at least 100 bp, more preferably at least 200bp, more preferably at least 300bp, more preferably at least 500bp, more preferably at least 600bp, more preferably at least 700bp, more preferably at least 800bp, more preferably at least 900bp, more preferably at least 1000bp, more preferably at least 1200bp, more preferably at least 1300bp and even more preferably at least 1400bp. Preferably, the functional fragment has a match percentage, i.e. positional identity of at least about 50%, more preferably of at least about 60%, even more preferably of at least about 70%, even more preferably of at least about 80%, even more preferably of at least about 85%, even more preferably of at least about 90%, even more preferably of at least about 95%, even more preferably of at least about 97%, even more preferably of at least about 98%, even more preferably of at least about 99% identity, and most preferably, the match percentage i.e. identity is equal to 100%.

Examples of polynucleotides are, but are not limited to, ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) molecules, either single stranded or double stranded. The person skilled in the art will comprehend that the complementary and reverse-complementary of the polynucleotides claimed herein are intended to fall within the scope of the present invention.

According to an embodiment of the invention, a sequence used for targeting to the adapted DNA domain is preferably a genetic element for introduction and / or expression of a polynucleotide of interest, wherein the functionality of the genetic element in the adapted DNA domain is disrupted upon integration of a construct. According to another embodiment of the invention, a sequence used for targeting to the adapted DNA domain is preferably a genetic element for introduction and / or expression of a polynucleotide of interest, wherein the functionality of the genetic element in the adapted DNA domain is restored upon integration of a construct.

It has previously been demonstrated that substantially homologous DNA domains can multiply through gene conversion and / or through gene amplification. Although being different, for use herein both terms gene conversion and gene amplification are used interchangeably. Different substantially homologous DNA domains revealed to have different frequencies of multiplication, i.e. gene conversion frequencies. The advantage of adapting a targeting DNA domain of a substantially homologous DNA domain with the highest gene conversion frequency is *inter alia* that screening and selection of a host with increased copy numbers of a polynucleotide of interest through gene conversion and / or amplification is much more efficient.

Accordingly, the substantially homologous DNA domain where the adapted substantially homologous DNA domain originates from has a gene conversion frequency that is preferably at least 10% higher than one of the other of the at least two substantially homologous DNA domains. More preferably, the gene conversion frequency is at least 15% higher, even more preferably at least 20% higher, even more preferably at least 30% higher, even more preferably at least 40% higher, even more preferably at least 50% higher, even more preferably at least 60% higher, even more preferably at least 70% higher, even more preferably at least 80% higher, even more preferably at least 90% higher, even more preferably at least 100% higher, even more preferably at least 150% higher, even more preferably at least 200% higher, even more preferably at least 300% higher, even more preferably at least 400% higher, even more preferably at least 500% higher, and even more preferably at least 1000% higher.

According to the invention a substantially homologous DNA domain having a higher gene conversion frequency can be identified by determining the gene conversion frequency of the substantially homologous DNA domains of a host cell. The gene conversion frequency can be determined by any means known to the person skilled in the art. A preferred method is the method described herein in example 3.1. In this method, a host cell comprising three substantially homologous DNA domains is transformed with a polynucleotide of interest. Three transformants are selected, each comprising a single copy of the polynucleotide of interest in a single substantially homologous DNA domain, each transformant having the polynucleotide integrated in a different substantially homologous DNA domain. After propagation of the transformants into offspring clones, the genotype of the offspring clones is determined by means of the unique sequence tag for identification purposes, i.e. using the "flag DNA" analysis described here above. The gene conversion frequency of a substantially homologous DNA domain is now calculated by dividing the amount of gene convertants of a transformant comprising said substantially homologous DNA domain by the total number of clones of said specific transformant.

Preferably, the host cell according to the invention is a cell wherein at least two of the substantially homologous DNA domains each have at least one copy of a polynucleotide of interest integrated. More preferably, each of said domains has at least two copies of the polynucleotide of interest integrated, even more preferably at least three copies, even more preferably four copies, even more preferably four to six copies. Preferably, each substantially homologous DNA domain of the host cell according to the invention has the same number of the polynucleotide of interest integrated.

The copy number of the polynucleotide of interest can be determined by any method available to the person skilled in the art, e.g. analysis of the genotype by Southern analysis or (semi)quantitative PCR, expression analysis by Northern analysis, GeneChip or microarray analysis or (semi)quantitative RT-PCR, analysis of the product encoded by the polynucleotide of interest by Western blot, ELISA, enzyme activity assay. Methods to determine copy numbers of the polynucleotide of interest are extensively described in WO9846772 and are according to the invention preferred methods to determine the copy number of the polynucleotide of interest.

Preferably, the host cell according to the invention comprises at least two substantially homologous DNA domains that are adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, more preferably three adapted DNA domains, even more preferably four adapted DNA domains, even more preferably five adapted DNA domains, even more preferably six adapted DNA domains. Most preferably, each substantially homologous DNA domain of said host cell according to the invention is an adapted DNA domain. Preferably, an adapted DNA domain has at least two copies of the polynucleotide of interest integrated, even more preferably at least three copies, even more preferably at least four copies, even more preferably an adapted DNA domain has at least five copies, even more preferably at least six copies, at least seven copies, at least eight copies, at least nine copies, at least ten copies of the polynucleotide of interest integrated. Most preferably, an adapted DNA domains has five or six copies of the polynucleotide of interest integrated. Preferably, each of said adapted DNA domains has the same number of the polynucleotide of interest integrated.

Preferably, said host cell comprising multiple substantially homologous DNA domains that are adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, wherein preferably each adapted DNA domain has multiple copies of the polynucleotide of interest integrated, is derived from a host cell comprising a single adapted DNA domain which has multiple copies of the polynucleotide of interest integrated; i.e. the additional copies of the adapted DNA domain which has multiple copies of the polynucleotide of interest integrated are generated by gene conversion.

Preferably, the efficiency of targeted integration of a polynucleotide to a pre-determined site into the genome of the host cell according to the invention is increased by steering an integration pathway towards HR. Such steering can be achieved either by elevating the efficiency of the HR pathway, and/or by lowering (meaning reducing) the efficiency of the NHR pathway and/or by decreasing the NHR/HR ratio.

Eukaryotic cells have at least two separate pathways (one via homologous (HR) and one via non-homologous recombination (NHR)) through which polynucleotides can be integrated into the host genome. In the context of the invention, the HR pathway is

defined as all genes and elements being involved in the control of the targeted integration of polynucleotides into the genome of a host, said polynucleotides having a certain homology with a certain pre-determined site of the genome of a host wherein the integration is targeted. The NHR pathway is defined as all genes and elements being involved in the control of the integration of polynucleotides into the genome of a host, irrespective of the degree of homology of the said polynucleotides with the genome sequence of the host.

The ratio of non-homologous to homologous recombination (NHR/HR) determines the efficiency of targeted integration of a polynucleotide to a pre-determined site in the genome of the host cell. When the ratio NHR/HR is high, the efficiency of targeted integration will be relatively low. When the ratio NHR/HR is low, the efficiency of targeted integration will be relatively high. The ratio of NHR/HR can be determined by methods known by the person skilled in the art by means of assessing the frequencies of NHR and HR and subsequently dividing the respective frequencies, an example of a method is the method described in WO 02/052026. Preferably, the host cell according to the invention comprises a polynucleotide encoding an NHR component comprising a modification, wherein said host cell is deficient in the production of said NHR component compared to the parent cell it originates from when cultivated under comparable conditions. Said modification results in a decrease of the NHR/HR ratio, therewith increasing the efficiency of targeted integration of a polynucleotide to a pre-determined site in the genome of the host cell. By modification of a polynucleotide encoding an NHR component, resulting in a deficiency of said NHR component, the ratio NHR/HR of the host cell is steered towards HR and consequently, the efficiency of targeted integration will increase.

Preferably, the host cell according to the invention demonstrates at least 5% deficiency of said NHR component, more preferably at least 10% deficiency, more preferably at least 20% deficiency, more preferably at least 30% deficiency, more preferably at least 40% deficiency, more preferably at least 50% deficiency, more preferably at least 60% deficiency, more preferably at least 70% deficiency, more preferably at least 80% deficiency, more preferably at least 90% deficiency, more preferably at least 95% deficiency, more preferably at least 97% deficiency, more preferably at least 99% deficiency. Most preferably, the host cell demonstrates 100% deficiency of said NHR component.

The NHR component to be modified can be any NHR component known to the person skilled in the art. Preferred NHR components to be modified are selected from the group of: KU70, KU80, MRE11, RAD50, RAD51, RAD52, XRS2, SIR4, LIG4, or homologues thereof. More preferred NHR components to be modified are KU70 and KU80, *hdf*A and *hdf*B or homologues thereof. The most preferred NHR component to be modified is KU70 or *hdf*A, or a homologue thereof. Preferably, the host cell according to the invention comprises a polynucleotide encoding an NHR component comprising a disruption or deletion as described in WO2005/095624. Methods to obtain such host cell are known to the skilled person and are extensively described in WO2005/095624. Preferably, the ratio NHR/HR in the NHR deficient host cell according to the invention is decreased by at least 10 %, more preferably at least 20%, more preferably at least 30%, more preferably at least 40%, more preferably at least 50%, more preferably at least 100%, more preferably at least 200%, and even more preferably at least 1000%, as compared to the non-deficient parent host cell, when assayed under identical conditions.

Alternatively, or in combination with a deficiency of an NHR component, the host cell according to the invention has the expression level of at least one gene involved in HR up regulated compared to the expression level of the same gene in the cell where the host cell originates from. This can be achieved by increasing the expression level of a gene encoding a component involved in HR and/or by increasing the expression level of a component involved in HR and/or by (temporarily) increasing the activity of the component involved in HR.

Alternatively, or in combination with modulation of the NHR/HR ratio as described here above, the methods described in WO2007/115886 and WO2007/115887 may be used to positively select host cells wherein the polynucleotide of interest is integrated by homologous recombination and/or to de-select cells wherein integration has occurred by non-homologous recombination. WO2007/115886 and WO2007/115887 describe the use of (additional) selection markers for such selection or de-selection. An example of such selection is the selective killing by diphtheria toxin A of cells where integration has taken place by non-homologous recombination (WO2007/115887).

Modification of a polynucleotide is herein defined as any event resulting in a change in the sequence of the polynucleotide. A modification is construed as one or

more modifications. Modification may be accomplished by the introduction (insertion), substitution or removal (deletion) of one or more nucleotides in the nucleotide sequence or a regulatory element required for the transcription or translation of the polynucleotide. For example, nucleotides may be inserted or removed so as to result in the introduction of a stop codon, the removal of a start codon or a change or a frame-shift of the open reading frame of the polynucleotide. The modification of a sequence or a regulatory element thereof may be accomplished by site-directed or random mutagenesis, DNA shuffling methods, DNA reassembly methods, gene synthesis (see for example Young and Dong, (2004), Nucleic Acids Research 32, (7) electronic access http://nar.oupjournals.org/cgi/reprint/32/7/e59 or Gupta et al. (1968), Proc. Natl. Acad. Sci USA, 60: 1338-1344; Scarpulla et al. (1982), Anal. Biochem. 121: 356-365; Stemmer et al. (1995), Gene 164: 49-53), or PCR generated mutagenesis in accordance with methods known in the art. Examples of random mutagenesis procedures are well known in the art, such as for example chemical (NTG for example) mutagenesis or physical (UV for example) mutagenesis. Examples of directed mutagenesis procedures are the QuickChange™ site-directed mutagenesis kit (Stratagene Cloning Systems, La Jolla, CA), the 'The Altered Sites® II in vitro Mutagenesis Systems' (Promega Corporation) or by overlap extension using PCR as described in Gene. 1989 Apr 15;77(1):51-9. (Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR "Site-directed mutagenesis by overlap extension using the polymerase chain reaction") or using PCR as described in *Molecular Biology: Current Innovations and Future Trends.* (Eds. A.M. Griffin and H.G.Griffin. ISBN 1-898486-01-8;1995 Horizon Scientific Press, PO Box 1, Wymondham, Norfolk, U.K.).

Preferred methods of modification are based on techniques of gene replacement, gene deletion, or gene disruption. For example, in the gene disruption method, a polynucleotide corresponding to the endogenous polynucleotide is mutagenized in vitro to produce a defective polynucleotide which is then transformed into the parent cell to produce a defective polynucleotide. By homologous recombination, the defective polynucleotide replaces the endogenous polynucleotide. It may be desirable that the defective polynucleotide also encodes a marker, which may be used for selection of transformants in which the nucleic acid sequence has been modified. In cases of deletion or replacement of the endogenous polynucleotide, an appropriate DNA sequence has to be introduced at the target locus to be deleted or

replaced. The appropriate DNA sequence is preferably present on a cloning vector. Preferred integrative cloning vector comprises a DNA fragment, which is homologous to the polynucleotide to be deleted or replaced for targeting the integration of the cloning vector to this pre-determined locus. In order to promote targeted integration, the cloning vector is preferably linearized prior to transformation of the host cell. Preferably, linearization is performed such that at least one but preferably either end of the cloning vector is flanked by sequences homologous to the DNA sequence to be deleted or replaced.

Alternatively or in combination with other mentioned techniques, a technique based on *in vivo* recombination of cosmids in *E. coli* can be used, as described in: A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans* (2000) Chaveroche, M-K., Ghico, J-M. and d'Enfert C; Nucleic acids Research, vol 28, no 22.

Alternatively, modification, wherein said host cell is deficient in the production of a polynucleotide may be performed by established anti-sense techniques using a nucleotide sequence complementary to the nucleic acid sequence of the polynucleotide. More specifically, expression of the polynucleotide by a host cell may be reduced or eliminated by introducing a nucleotide sequence complementary to the nucleic acid sequence of the polynucleotide, which may be transcribed in the cell and is capable of hybridizing to the mRNA produced in the cell. Under conditions allowing the complementary anti-sense nucleotide sequence to hybridize to the mRNA, the amount of protein translated is thus reduced or eliminated. An example of expressing an antisense-RNA is shown in Appl Environ Microbiol. 2000 Feb;66(2):775-82. (Characterization of a foldase, protein disulfide isomerase A, in the protein secretory pathway of *Aspergillus niger*. Ngiam C, Jeenes DJ, Punt PJ, Van Den Hondel CA, Archer DB) or (Zrenner R, Willmitzer L, Sonnewald U. Analysis of the expression of potato uridinediphosphate-glucose pyrophosphorylase and its inhibition by antisense RNA. Planta. (1993);190(2):247-52.).

Furthermore, modification, downregulation or inactivation of the gene may be obtained via the RNA interference (RNAi) technique (FEMS Microb. Lett. 237 (2004): 317-324). In this method identical sense and antisense parts of the nucleotide sequence, which expression is to be affected, are cloned behind each other with a nucleotide spacer in between, and inserted into an expression vector. After such a molecule is transcribed, formation of small nucleotide fragments will lead to a targeted

degradation of the mRNA, which is to be affected. The elimination of the specific mRNA can be to various extends. The RNA interference techniques described in WO2008/053019, WO2005/05672A1, WO2005/026356A1, Oliveira et al., "Efficient cloning system for construction of gene silencing vectors in Aspergillus niger" (2008) Appl Microbiol and Biotechnol 80 (5): 917-924 and/or Barnes et al., "siRNA as a molecular tool for use in Aspergillus niger" (2008) Biotechnology Letters 30 (5): 885-890 may be used for downregulation, modification or inactivation of the gene.

Following modification of a polynucleotide, the obtained strains are screened for deficiency of the product encoded by the polynucleotide. The down and/or up regulation of the expression level of a polynucleotide can be monitored by quantifying the amount of corresponding mRNA present in a cell by Northern analysis (Sambrook and Russell (2001) "Molecular Cloning: A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, New York.) for example and/or by quantifying the amount of corresponding protein present in a cell by western blotting or ELISA, for example. The difference in mRNA amount may also be quantified by DNA array analysis (Eisen, M.B. and Brown, P.O. DNA arrays for analysis of gene expression. Methods Enzymol. 1999:303:179-205).

Deficiency of a host cell is herein defined as a phenotypic feature wherein the cell produces less of the product encoded by the modified polynucleotide and/or has a reduced expression level of a modified polynucleotide (i.e. reduced mRNA level) and/or has a decreased specific (protein) activity of the product encoded by the modified polynucleotide and combinations of these possibilities as compared to the parent cell comprising the un-modified polynucleotide.

According to the invention, the substantially homologous DNA domain is a DNA domain or locus (terms used interchangeably herein) which in its native state comprises an endogenous gene capable of high level expression. The term "endogenous" gene is herein defined as a naturally occurring copy of a gene in the genome of the organism in question.

It is generally known that the expression level of an integrated recombinant gene can vary greatly depending on the genomic locus where that gene is integrated. The advantage of using highly expressed DNA domains for integration of recombinant

genes to be expressed is that these DNA domains are at least capable of supporting high level expression of the endogenous gene. It is therefore likely that such DNA domains will also support high level expression of an integrated recombinant gene. Previously, it was determined that integration in the *gla*A locus of *A.niger* and integration in the penicillin cluster of *P.chrysogenum* provides higher expression levels per gene copy as compared to integration in some other genomic loci (see WO9846772). In this context it will be understood that a gene capable of high level expression is defined as a gene which, when expressed at maximum level, produces an mRNA that constitutes at least 0.1 % of the total mRNA population, preferably at least 0.5 % of the total mRNA and most preferably at least 1 % of the total mRNA. Examples of such highly expressible endogenous genes of which the DNA domains in which they are contained are particularly suitable as substantially homologous DNA domains according to the invention are genes encoding glycolytic enzymes, amylolytic enzymes, cellulolytic enzymes and/or antibiotic biosynthetic enzymes. Even more preferred are loci comprising genes involved in industrial processes and known to be expressed at high level such as glucoamylase genes, TAKA amylase genes, cellobiohydrolase genes and penicillin biosynthetic genes.

Preferably, the at least two substantially homologous DNA domains of the host cell according to the invention are loci of glaA or are loci of amyA or are loci of amyB or are fragments or homologues of these loci. More preferred loci are the glaA, amyA and amyB loci of *Aspergillus*. Even more preferred loci are the glaA, and amyB loci of *Aspergillus niger.* Even more preferred loci are the glaA, amyBI and amyBII loci of *Aspergillus niger* CBS 513.88 (gene naming as in WO2005/095624). Even more preferred loci are the glaA, amyBI and amyBII loci (nucleotide sequence of glaA gene (An03g06550) and its genomic context a.o. can be derived from http://www.ncbi.nlm.nih.gov/; idem for amyBI gene sequence (An12g06930) and amyBII gene sequence (An05g02100) or are fragments or homologues of these loci.

For purposes of the invention, the terms “homology” and “identity” are used interchangeably. The degree of homology (identity) between two nucleic acid sequences is herein preferably determined by the BLAST program. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/).The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The

BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

According to the invention, the highly expressed endogenous gene is preferably inactivated in each copy of the DNA domain in the host cell according to the invention in cases where the expression of the endogenous gene is not required. In such cases the inactivation of the high level expression of the endogenous gene releases energy and resources which can further be utilized for the expression of the polynucleotide of interest. Moreover, in case both a desired compound to be produced by integration of the polynucleotide of interest and the enzyme encoded by the endogenous gene are secreted enzymes, inactivation of the endogenous enzyme will result in more pure preparations of the desired enzyme. Preferably the endogenous gene is inactivated by means of an irreversible deletion of at least part of the endogenous gene in order to exclude reversion of the inactivation. More preferably the inactivation of the endogenous gene is effected by an irreversible deletion which comprises at least part of the promoter and upstream activating sequences. This is particularly advantageous in cases where the expression of a desired gene encoding an enzyme to be produced by integration of the recombinant DNA molecule is driven from a promoter derived from the endogenous gene because it will eliminate competition for potentially limiting transcription factors required for expression of the desired gene.

The polynucleotide of interest may be any polynucleotide. The polynucleotide of interest may be obtained from any prokaryotic, eukaryotic, or other source. Preferably, the polynucleotide of interest and promoter associated with it are homologous to the host cell, resulting in a recombinant host cell being a self-clone.

According to the invention, the polynucleotide of interest may be a variant, optimized polynucleotide comprising an optimized terminator sequence, such as for example described in WO2006 077258. The polynucleotide of interest may be a partly synthetic polynucleotide or an entirely synthetic nucleic acid sequence. The polynucleotide of interest may be optimized in its codon use, preferably according to the methods described in WO2006/077258 and/or WO2008/000632, which are herein incorporated by reference. WO2008/000632 addresses codon-pair optimization.

Codon-pair optimisation is a method wherein the nucleotide sequences encoding a polypeptide have been modified with respect to their codon-usage, in particular the codon-pairs that are used, to obtain improved expression of the nucleotide sequence encoding the polypeptide and/or improved production of the encoded polypeptide. Codon pairs are defined as a set of two subsequent triplets (codons) in a coding sequence.

Accordingly, according to the invention, the polynucleotide of interest is preferably a codon optimized polynucleotide.

The polynucleotide preferably comprises a coding sequence and may be comprised in a nucleic acid construct and/or in a vector, further comprising a promoter and control sequences to facilitate, cloning, transformation, expression and/or production of the compound encoded by the polynucleotide of interest.

"Nucleic acid construct " is defined herein as a nucleic acid molecule, either single-or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acid which are combined and juxtaposed in a manner which would not otherwise exist in nature. The term nucleic acid construct is synonymous with the term “expression cassette” when the nucleic acid construct contains all the control sequences required for expression of a coding sequence. The term "coding sequence" as defined herein is a sequence, which is transcribed into mRNA and translated into a polypeptide. The boundaries of the coding sequence are generally determined by the ATG start codon at the 5' end of the mRNA and a translation stop codon sequence terminating the open reading frame at the 3'end of the mRNA. A coding sequence can include, but is not limited to, DNA, cDNA, and recombinant nucleic acid sequences.

The term "control sequences" is defined herein to include all components, which are necessary or advantageous for the expression of a polypeptide. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, optimal translation initiation sequences (as described in Kozak, 1991, J. Biol. Chem. 266:19867-19870), a polyadenylation sequence, a pro-peptide sequence, a pre-pro-peptide sequence, a promoter, a signal sequence, and a transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals.

The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide. The term "operably linked" is defined herein as a configuration in which a control sequence is appropriately placed at a position relative to the coding sequence of the DNA sequence such that the control sequence directs the production of a polypeptide.

The control sequence may be an appropriate promoter sequence, a nucleic acid sequence, which is recognized by a host cell for expression of the nucleic acid sequence. The promoter sequence contains transcriptional control sequences, which mediate the expression of the polypeptide. The promoter may be any nucleic acid sequence, which shows transcriptional activity in the cell including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

As promoter will also be understood the 5' non-coding region (between promoter and translation start) for translation after transcription into mRNA, cis-acting transcription control elements such as enhancers, and other nucleotide sequences capable of interacting with transcription factors.

The promoter may be the promoter natively associated with the coding sequence to be expressed. The promoter may also be a constitutive or inducible promoter foreign to the coding sequence to be expressed. Examples of suitable promoters for use in mammalian cells are e.g. described in Sambrook and Russell, *supra*. Examples of suitable promoters for use in yeasts include e.g. glycolytic promoters.

Examples of preferred inducible promoters that can be used are a starch-, copper-, oleic acid- inducible promoters.

Preferably, the promoter is selected from the group, which includes but is not limited to promoters obtained from the genes encoding *A. oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *A. niger* neutral alpha-amylase, *A. niger* acid stable alpha-amylase, *A. niger* or *A. awamori* glucoamylase (*gla*A), *R. miehei* lipase, *A. oryzae* alkaline protease, *A. oryzae* triose phosphate isomerase, *A. nidulans* acetamidase, the NA2-tpi promoter (a hybrid of the promoters from the genes encoding *A. niger* neutral alpha-amylase and *A. oryzae* triose phosphate isomerase, WO03/008575), and mutant, truncated, and hybrid promoters thereof. Particularly

preferred promoters for use in filamentous fungal cells are a promoter, or a functional part thereof, from a protease gene; e. g., from the *F. oxysporum* trypsin-like protease gene (U. S. 4, 288, 627), *A. oryzae* alkaline protease gene(*alp*), *A. niger pac*A gene, *A. oryzae* alkaline protease gene, *A. oryzae* neutral metalloprotease gene, *A. niger* aspergillopepsin protease *pep*A gene, or *F. venenatum* trypsin gene, *A. niger* aspartic protease *pep*B gene.

A single appropriate promoter sequence may be used as control sequence for expression of the polynucleotide of interest, or multiple distinct promoter sequences may be used as control sequences for expression of the polynucleotide of interest. If multiple distinct promoter sequences are used, preferably at least two distinct promoter sequences are each operably linked to a polynucleotide of interest, according to WO2008/098933.

The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a filamentous fungal cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator, which is functional in the cell, may be used in the present invention.

Preferred terminators for filamentous fungal cells are obtained from the genes encoding *A. oryzae* TAKA amylase, *A. niger* glucoamylase (*gla*A), *A. nidulans* anthranilate synthase, *A. niger* alpha-glucosidase, trpC gene and *Fusarium oxysporum* trypsin-like protease.

The control sequence may also be a suitable leader sequence, a non-translated region of a mRNA which is important for translation by the filamentous fungal cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence, which is functional in the cell, may be used in the present invention.

Preferred leaders for filamentous fungal cells are obtained from the genes encoding *A. oryzae* TAKA amylase and *A. nidulans* triose phosphate isomerase and *A. niger gla*A.

Other control sequences may be isolated from the Penicillium IPNS gene, or *pcb*C gene, the beta tubulin gene. All the control sequences cited in WO 01/21779 are contemplated to be envisioned for use in the present invention.

The control sequence may also be a polyadenylation sequence, a sequence which is operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the filamentous fungal cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence, which is functional in the cell, may be used in the present invention.

Preferred polyadenylation sequences for filamentous fungal cells are obtained from the genes encoding *A. oryzae* TAKA amylase, *A. niger* glucoamylase, *A. nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease and *A. niger* alpha-glucosidase.

Preferably, the DNA construct comprises the polynucleotide of interest encoding a compound of interest, a promoter DNA sequence as described here above, and preferred control sequences such as:

- one translational termination sequence orientated in 5' towards 3' direction selected from the following list of sequences: TAAG, TAGA and TAAA, preferably TAAA, and/or
- one translational initiator coding sequence orientated in 5' towards 3' direction selected from the following list of sequences: GCTACCCCC; GCTACCTCC; GCTACCCTC; GCTACCTTC; GCTCCCCCC; GCTCCCTCC; GCTCCCCTC; GCTCCCTTC; GCTGCCCCC; GCTGCCTCC; GCTGCCCTC; GCTGCCTTC; GCTTCCCCC; GCTTCCTCC; GCTTCCCTC; and GCTTCCTTC, preferably GCT TCC TTC, and/or
- one translational initiator sequence selected from the following list of sequences: 5'-mwChkyCAAA-3'; 5'-mwChkyCACA-3' or 5'-mwChkyCAAG-3', using ambiguity codes for nucleotides: m (A/C); w (A/T); y (C/T); k (G/T); h (A/C/T), preferably 5'-CACCGTCAAA-3' or 5'-CGCAGTCAAG-3'.

In the context of this invention, the term "translational initiator coding sequence" is defined as the nine nucleotides immediately downstream of the initiator or start codon of the open reading frame of a DNA coding sequence. The initiator or start

codon encodes for the AA methionine. The initiator codon is typically ATG, but may also be any functional start codon such as GTG.

In the context of this invention, the term "translational termination sequence" is defined as the four nucleotides starting from the translational stop codon at the 3' end of the open reading frame or nucleotide coding sequence and oriented in 5' towards 3' direction.

In the context of this invention, the term "translational initiator sequence" is defined as the ten nucleotides immediately upstream of the initiator or start codon of the open reading frame of a DNA sequence coding for a polypeptide. The initiator or start codon encodes for the AA methionine. The initiator codon is typically ATG, but may also be any functional start codon such as GTG. It is well known in the art that uracil, U, replaces the deoxynucleotide thymine, T, in RNA.

The polynucleotide of interest encoding a compound of interest, or DNA construct comprising the polynucleotide of interest and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of optional polynucleotide sequences.

Alternatively, the polynucleotide of interest may be expressed by inserting the sequence or a nucleic acid construct comprising polynucleotide of interest into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression, and possibly secretion.

The recombinant expression vector may be any vector (e.g., a plasmid or virus), which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the polynucleotide of interest. The choice of the vector will typically depend on the compatibility of the vector with host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids. The vector may be an autonomously replicating vector, i. e., a vector, which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. An autonomously maintained cloning vector may comprise the AMA1-sequence (see e.g. Aleksenko and Clutterbuck (1997), Fungal Genet. Biol. 21: 373-397).

Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome (s) into which it has been integrated. Preferably, the integrative cloning vector comprises a DNA fragment, which is homologous to a DNA sequence in a predetermined target locus in the genome of the host cell for targeting the integration of the cloning vector to this predetermined locus. In order to promote targeted integration, the cloning vector is preferably linearized prior to transformation of the host cell. Linearization is preferably performed such that at least one but preferably either end of the cloning vector is flanked by sequences homologous to the target locus. The length of the homologous sequences flanking the target locus is preferably at least 30bp, preferably at least 50 bp, preferably at least 0.1kb, even preferably at least 0.2kb, more preferably at least 0.5 kb, even more preferably at least 1 kb, most preferably at least 2 kb.

The vector system may be a single vector or plasmid or two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the filamentous fungal cell, or a transposon.

The vectors preferably contain one or more selectable markers, which permit easy selection of transformed cells. Using the method of co-transformation, one vector may contain the selectable marker whereas another vector may contain the polynucleotide of interest or the nucleic acid construct of interest; the vectors are simultaneously used for transformation of the host cell. The host cell according to the invention has the ability that co-transformation efficiency can be very high, up to 100% (see example 12.2).

A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. A selectable marker for use in a filamentous fungal host cell may be selected from the group including, but not limited to, *amd*S (acetamidase), *arg*B (ornithine carbamoyltransferase), *bar* (phosphinothricinacetyltransferase), *ble*A (phleomycin binding), *hyg*B (hygromycinphosphotransferase), *nia*D (nitrate reductase), *pyr*G (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenyltransferase), and *trp*C (anthranilate synthase), as well as equivalents from other species. Preferred for use in an *Aspergillus* and *Penicillium* cell are the *amd*S (EP 635574 B1, WO 97/06261) and pyrG genes of *A. nidulans* or *A. oryzae* and the *bar* gene of *Streptomyces*

hygroscopicus. More preferably an *amd*S gene is used, even more preferably an *amd*S gene from *A. nidulans* or *A. niger*. A most preferred selection marker gene is the *A.nidulans amdS* coding sequence fused to the *A.nidulans gpdA* promoter (see EP 635574 B1). *Amd*S genes from other filamentous fungi may also be used (WO 97/06261).

The procedures used to ligate the elements described above to construct the recombinant expression vectors of the present invention are well known to one skilled in the art (see, *e.g.*, Sambrook and Russell, *supra*; and Ausubel *et al.*, *Current Protocols in Molecular Biology*, Wiley InterScience, NY, 1995).

Preferably, the host cell according to the invention comprises, in addition to adapted DNA domains and optionally increased efficiency of targeted integration, a polynucleotide selected from the group of: glaA, amyA, amyBI, amyBII, oahA, toxin associated polynucleotide and prtT, said polynucleotide comprising a modification, wherein the host cell is deficient in the product encoded by the polynucleotide comprising the modification, compared to the parent cell it originates from when cultivated under comparable conditions. Modification is defined as earlier herein.

Preferably, the host cell according to the invention additionally comprises a modification of Sec61. Deficiency of a host cell is herein defined, as previously herein, as a phenotypic feature wherein the cell produces less of the product encoded by the modified polynucleotide and/or has reduced expression level of the modified polynucleotide (i.e. reduced mRNA level) and/or has decreased specific (protein) activity of the product encoded by the modified polynucleotide or combinations of these features as compared to the parent cell comprising the un-modified polynucleotide.

Preferably, the host cell according to the invention demonstrates at least 5% deficiency of at least one of glaA, amyA, amyBI, amyBII, oahA, toxin associated polynucleotide or prtT, more preferably at least 10% deficiency, more preferably at least 20% deficiency, more preferably at least 30% deficiency, more preferably at least 40% deficiency, more preferably at least 50% deficiency, more preferably at least 60% deficiency, more preferably at least 70% deficiency, more preferably at least 80% deficiency, more preferably at least 90% deficiency, more preferably at least 95% deficiency, more preferably at least 97% deficiency, more preferably at least 99%

deficiency. Most preferably, the host cell demonstrates 100% deficiency of at least one of glaA, amyA, amyBI, amyBII, oahA, toxin associated polynucleotide or prtT.

The advantage of the deficiency of one or more polypeptides selected from the group of glucoamylase (glaA), acid stable alpha-amylase (amyA), neutral alpha-amylase (amyBI and amyBII) and oxalic acid hydrolase (oahA) in the production of a compound of interest is that energy and resources are not utilized for these by-products. This energy and resources can be used for the production of a compound encoded by the polynucleotide of interest. Furthermore, downstream processing of a compound of interest is simplified since there are fewer by-products present.

Oxalic acid hydrolase (oahA) is a component of the synthesis pathway of oxalic acid in many host cells. A host cell deficient in oahA will be deficient in oxalic acid. Oxalic acid is an unwanted by-product in many applications such as food-applications. Furthermore, oxalic acid lowers the pH of the medium cultivations of host cell producing this component, resulting in lowered yields; i.e. yield is increased in oxalic acid deficient host cells. It is therefore advantageous if the host cell according to the invention is deficient in oahA. OahA deficient host cells and preferred methods of producing said host cells are extensively described in WO 2000/50576 and WO2004/070022. A preferred method to produce an oahA deficient host cell is the recombinant method of disruption described in WO 2000/50576. Preferably, the host cell according to the invention is deficient in oahA. Preferably, the oahA is a fungal oahA. More preferably, the oahA is the oahA from *Aspergillus*. Even more preferably the oahA is the oahA from *Aspergillus niger*. Even more preferably the oahA is the oahA from *Aspergillus niger* CBS 513.88. Most preferably, the oahA comprises the sequence of An10g00820.

prtT is a transcriptional activator of proteases in eukaryotic cells. Several fungal transcriptional activators of proteases have been recently described in WO 00/20596, WO 01/68864, WO 2006/040312 and WO 2007/062936. These transcriptional activators were isolated from *Aspergillus niger (A. niger), Aspergillus fumigatus (A. fumigatus), Penicillium chrysogenum (P. chrysogenum)* and *Aspergillus oryzae (A. oryzae)*. These transcriptional activators of protease genes can be used to improve a method for producing a polypeptide in a fungal cell, wherein the polypeptide is sensitive for protease degradation. When the host cell is deficient in prtT, the host cell will produce less proteases that are under transcriptional control of prtT. It is therefore

advantageous when the host cell according to the invention is deficient in prtT. prtT deficient hosts and preferred methods to produce these hosts are extensively described in WO 01/68864, WO 2006/040312. WO 01/68864 and WO 2006/040312 describe recombinant and classic methods to disrupt the prtT coding sequence. WO 2007/062936 describes disruption of the prtT binding site in a protease promoter. Disruption of the binding site impedes binding of prtT to the binding site. Consequently, the transcription of the protease is not activated by prtT and less protease is produced.

Preferably, the host cell according to the invention comprises a polynucleotide encoding prtT, said polynucleotide comprising a modification, wherein the host cell is deficient in the production of prtT compared to the parent cell it originates from when cultivated under comparable conditions. Preferably, the prtT is a fungal prtT. More preferably, the prtT is the prtT from *Aspergillus*. Even more preferably the prtT is the prtT from *Aspergillus niger*. Even more preferably the prtT is the prtT from *Aspergillus niger* CBS 513.88. Most preferably, the prtT comprises the sequence of An04g06940.

The term “glucoamylase” (glaA) is identical to the term "amyloglucosidase" and is defined herein as an enzyme having dextrin 6-alpha-D-glucanohydrolase activity which catalyses the endo hydrolysis of 1, 6-alpha-D-glucoside linkages at points of branching in chains of 1, 4-linked alpha-D-glucose residues and terminal 1, 4-linked alpha-D-glucose residues. Glucoamylase activity can be measured as AGIU/ml by determining the liberation of paranitrofenol from the substrate p-nitrophenyl-a-D-glucopyranoside (Sigma). This results in a yellow colour, whose absorbance can be measured at 405 nm using a spectrophotometer. 1 AGIU is the quantity of enzyme, which produces 1 µmole of glucose per minute at pH 4.3 and 60°C from a soluble starch substrate. In WO98/46772 additional details of the assay can be found.

Preferably, the host cell according to the invention comprises a polynucleotide encoding glaA, said polynucleotide comprising a modification, wherein the host cell is deficient in the production of glaA compared to the parent cell it originates from when cultivated under comparable conditions. Preferably, the glaA is a fungal glaA. More preferably, the glaA is the glaA from *Aspergillus*. Even more preferably the glaA is the glaA from *Aspergillus niger*. Even more preferably the glaA is the glaA from *Aspergillus niger* CBS 513.88. Most preferably, the glaA comprises the sequence of An03g06550.

The term "alpha-amylase" is defined herein as 1, 4-alpha-D-glucan glucanohydrolase activity which catalyzes the endohydrolysis of polysaccharides with three or more alpha-1, 4-linked glucose units in the presence of water to malto-oligosaccharides. To determine the (neutral) alpha-amylase activity, the Megazyme cereal alpha-amylase kit is used (Megazyme, CERALPHA alpha amylase assay kit, catalogus. ref. K-CERA, year 2000-2001), according a protocol of the supplier. The measured activity is based on hydrolysis of non-reducing-endblocked ρ-nitrophenyl maltoheptaoside in the presence of excess glucoamylase and α-glucosidase at a pH of 7.0. The amount of formed ρ-nitrophenol is a measure for alpha-amylase activity present in a sample.

The term "acid stable alpha-amylase" (amyA) is defined herein as an enzyme having alpha-amylase activity with optimal activity in the acid pH range. To determine the acid stable alpha-amylase activity, also the Megazyme cereal alpha-amylase kit is used (Megazyme, CERALPHA alpha amylase assay kit, catalogus. ref. K-CERA, year 2000-2001), according a protocol of the supplier but at an acid pH. The measured activity is based on hydrolysis of non-reducing-endblocked ρ-nitrophenyl maltoheptaoside in the presence of excess glucoamylase and α-glucosidase at a pH of 4.5. The amount of formed ρ-nitrophenol is a measure for acid stable alpha-amylase activity present in a sample.

Preferably, the host cell according to the invention comprises a polynucleotide encoding AmyA, said polynucleotide comprising a modification, wherein the host cell is deficient in amyA compared to the parent cell it originates from when cultivated under comparable conditions. Preferably, the amyA is a fungal amyA. More preferably, the amyA is the amyA from *Aspergillus*. Even more preferably the amyA is the amyA from *Aspergillus niger*. Even more preferably the amyA is the amyA from *Aspergillus niger* CBS 513.88. Most preferably, the amyA comprises the sequence of An11g03340.

The term "neutral alpha-amylase activity" (amy) is defined herein as an enzyme having alpha-amylase activity with optimal activity in the neutral pH range.

Preferably, the host cell according to the invention comprises a polynucleotide encoding AmyB, said polynucleotide comprising a modification, wherein the host cell is

deficient in amyBI and/or amyBII compared to the parent cell it originates from when cultivated under comparable conditions. More preferably, the host cell is deficient in amyBI and amy BII. Preferably, the amyB a is a fungal amyB. More preferably, the amyB is the amyB from *Aspergillus*. Even more preferably the amyB is the amyBI from *Aspergillus niger*. Even more preferably the amyB is the amyBI from *Aspergillus niger* CBS 513.88. Most preferably, the amyBI comprises the sequence of An12g06930. Even more preferably the amyB is the amyBII from *Aspergillus niger*. Even more preferably the amyB is the amyBII from *Aspergillus niger* CBS 513.88. Most preferably, the amyBII comprises the sequence of An05g02100.

The term toxin associated polynucleotide is defined herein as a gene cluster, a multitude of genes, a gene or part thereof encoding a compound, or biochemical pathway responsible for the biosynthesis or secretion of at least one toxin or toxin intermediate compound. Said compound may e.g. be a polypeptide, which may be an enzyme.

A number of host cells, especially fungi, which are used as host cells in the production of polypeptides of interest possesses genes encoding enzymes involved in the biosynthesis of various toxins. For example, cyclopiazonic acid, kojic acid, 3-nitropropionic acid and aflatoxins are known toxins, which are formed in, e.g., *Aspergillus flavus*. Similarly, trichothecenes are formed in a number of fungi, e.g., in *Fusarium sp*. such as *Fusarium venenatum* and in *Trichoderma* and ochratoxin may be produced by *Aspergillus*. Recently, sequencing of the genome of an industrial *Aspergillus niger* host strain revealed a fumonisin gene cluster (Pel et al., “Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88”. Nat Biotechnol. 2007 Feb; 25 (2):221-231). The formation of such toxins during the fermentation of compounds of interest is highly undesirable as these toxins may present a health hazard to operators, customers and the environment. Consequently, a toxin deficient host cell enables toxin-free production of a compound of interest. The toxin-free compound is easier to produce since no toxin has to be removed from the product. Furthermore, the regulatory approval procedure for the compound is easier.

Preferably, the host cell according to the invention comprises a toxin associated polynucleotide encoding a compound (which may e.g. be a polypeptide which may be an enzyme) or biochemical pathway, said toxin associated polynucleotide comprising a

modification, wherein the host cell is deficient in the production of said toxin or a toxin intermediate compound compared to the parent cell it originates from when cultivated under comparable conditions. Preferably, the toxin or toxin intermediate compound is a fungal toxin or toxin intermediate compound. More preferably, the toxin or toxin intermediate compound is a toxin or toxin intermediate compound from *Aspergillus*. Even more preferably the toxin or the toxin intermediate compound is a toxin or toxin intermediate compound from *Aspergillus niger*. Even more preferably the toxin or toxin intermediate compound is a toxin or toxin intermediate compound from *Aspergillus niger* CBS 513.88. Even more preferably, the toxin or the toxin intermediate compound is fumonisin or a fumonisin intermediate compound. Even more preferably, the toxin or the toxin intermediate compound is ochratoxin or an ochratoxin intermediate compound. Most preferably, the toxin or the toxin intermediate compound is ochratoxin or fumonisin or an ochratoxin or a fumonisin intermediate compound.

Preferably, the toxin associated polynucleotide encodes a compound (which may e.g. be a polypeptide which may be an enzyme) or a biochemical pathway which is involved in the production of a fungal toxin or toxin intermediate compound. More preferably, a toxin or toxin intermediate compound from *Aspergillus*. Even more preferably, a toxin or toxin intermediate compound from *Aspergillus niger*. Even more preferably, a toxin or toxin intermediate compound from *Aspergillus niger* CBS 513.88. Even more preferably, a fumonisin or a fumonisin intermediate compound. Even more preferably, a fumonisin-B or a fumonisin-B intermediate compound. Even more preferably, a fumonisin-B2 or a fumonisin-B2 intermediate compound. Even more preferably, the toxin associated polynucleotide comprises the sequence of the fumonisin cluster from An01g06820 until An01g06930. Most preferably, the toxin associated polynucleotide comprises the sequence of An01g06930.

In another preferred embodiment, the toxin associated polynucleotide encodes a compound (which may e.g. be a polypeptide which may be an enzyme) or a biochemical pathway which is involved in ochratoxin or an ochratoxin intermediate compound. More preferably, an ochratoxin A or an ochratoxin A intermediate compound. More preferably, the toxin associated polynucleotide comprises the sequence of the cluster from An15g07880 until An15g07930. Most preferably, the toxin associated polynucleotide comprises the sequence of An15g07910 and/or the sequence of An15g07920.

Preferably, the host cell according to the invention comprises at least one toxin associated polynucleotide encoding a compound (which may e.g. be a polypeptide which may be an enzyme) or biochemical pathway, said toxin associated polynucleotide comprising at least one modification, wherein the host cell is deficient in the production of a toxin or, toxin intermediate compound compared to the parent cell it originates from when cultivated under comparable conditions.

More preferably, the host cell according to the invention comprises two toxin associated polynucleotides, said two toxin associated polynucleotides each comprising at least one modification, wherein the host cell is preferably deficient in the production of fumonisin and ochratoxin compared to the parent cell it originates from when cultivated under comparable conditions.

Even more preferably, the host cell according to the invention comprises three or more toxin associated polynucleotides said three or more toxin associated polynucleotides each comprising at least one modification, wherein the host cell is preferably deficient in the production of fumonisin, ochratoxin and at least one additional toxin or toxin intermediate compound compared to the parent cell it originates from when cultivated under comparable conditions.

Preferably, the host cell according to the invention is deficient in glaA and at least one of the components selected from the group of amyA, amyBI, amyBII, oahA, toxin associated compound and prtT, by virtue of having a modification in the polynucleotide encoding glaA and said components. More preferably, the host cell according to the invention is deficient in glaA, oahA and at least one of the components selected from the group of amyA, amyBI, amyBII, toxin associated compound and prtT Even more preferably, the host cell according to the invention is deficient in glaA, oahA, toxin associated compound and at least one of the components selected from the group of amyA, amyBI, amyBII, and prtT. Even more preferably the host cell according to the invention is deficient in glaA, oahA, amyA, amyBI, amyBII and at least one of the components selected from the group of toxin associated compound and prtT. Even more preferably the host cell according to the invention is deficient in glaA, oahA, amyA, amyBI, amyBII, prtT and toxin associated compound. Most preferably, the host cell according to the invention comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest

wherein at least one of the at least two substantially homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, is deficient in glaA, oahA, amyA, amyBI, amyBII, prtT and toxin associated compound and is furthermore deficient in an NHR component, preferably ku70 or a homologue thereof.

In addition to deficiency or modification of the components described here above, the host cell according to the invention may be deficient in other components, such as major proteases like pepA. Preferably, the pepA is a fungal pepA. More preferably, the pepA is the pepA from *Aspergillus*. Even more preferably the pepA is the pepA from *Aspergillus niger*. Even more preferably the pepA is the pepA from *Aspergillus niger* CBS 513.88. Most preferably, the pepA comprises the sequence of An14g04710. Preferably, the host cell according to the invention demonstrates at least 5% deficiency of *pepA*, more preferably at least 10% deficiency, more preferably at least 20% deficiency, more preferably at least 30% deficiency, more preferably at least 40% deficiency, more preferably at least 50% deficiency, more preferably at least 60% deficiency, more preferably at least 70% deficiency, more preferably at least 80% deficiency, more preferably at least 90% deficiency, more preferably at least 95% deficiency, more preferably at least 97% deficiency, more preferably at least 99% deficiency. Most preferably, the host cell demonstrates 100% deficiency of *pepA*.

Alternatively, or in combination with the deficiencies described here above, the host cell comprises an elevated unfolded protein response (UPR) compared to the wild type cell to enhance production abilities of a polypeptide of interest. UPR may be increased by techniques described in US2004/0186070A1 and/or US2001/0034045A1 and/or WO01/72783A2 and/or WO2005/123763. More specifically, the protein level of HAC1 and/or IRE1 and/or PTC2 has been modulated, and/or the SEC61 protein has been engineered in order to obtain a host cell having an elevated UPR. A preferred SEC61 modification is a modification which results in a one-way mutant of SEC61; i.e. a mutant wherein the *de novo* synthesized protein can enter the ER via SEC61, but the protein cannot leave the ER via SEC61. Such modifications are extensively described in WO2005/123763. Most preferably, the SEC 61 modification is the S376W mutation in which Serine 376 is replaced by Tryptophan.

A preferred host cell according to the invention is a recombinant *Aspergillus niger* recombinant host cell for the production of a compound of interest, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest wherein at least one of the at least two substantially homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, and wherein the substantially homologous DNA domain where the adapted substantially homologous DNA domain originates from has a gene conversion frequency that is at least 10% higher than one of the other of the at least two substantially homologous DNA domains, wherein the adapted substantially homologous DNA domain is distinguished from the other versions of the substantially homologous DNA domains by means of a unique sequence tag and wherein the adapted substantially homologous DNA domain comprises a targeting DNA domain, wherein said targeting DNA domain comprises a sequence with enhanced integration preference and said host cell being deficient in glaA, pepA, hdfA, amyBII, amyBI, amyA, oahA, fumB, och. Preferably, said host cell has a sec61 S376W mutation, i.e. wherein the S376W mutation in which Serine 376 is replaced by Tryptophan. More preferably, said host cell additionally is deficient in prtT. Such host cell are described in examples 9, 10 and 11.

Preferably, the host cell according to the invention is a selection marker free host cell.

This situation may be obtained by counter-selection for a bidirectional dominant selection marker, such as the acetamidase gene (amdS) from *Aspergillus nidulans*. This marker can be used to select for transformants having the gene by selecting with acetamide as sole carbon and/or nitrogen source, whereas counter-selection (a term reserved hereinafter for selection for the absence of the marker gene) can be done with, for example fluoracetamide. Other bidirectional dominant selection markers which work in specific host cells can be used in an analogous fashion. The out-recombination of the bidirectional marker gene is facilitated by inserting the gene flanked by direct repeats on the incoming plasmid.

As disclosed in European patent EP0635574B1, the amdS bidirectional marker is dominant in both directions, meaning that transformed cells of any genetic

background can be selected for the presence of the marker (using acetamide as sole carbon and/or sole nitrogen source). Other bidirectional markers are URA3, LYS2, pyrG, facA and the like.

The use of cells which are marker-free is more easily accepted by the regulatory authorities, and at the same time pose less burden to the energy balance of the yeast under industrial fermentation conditions.

The host cell according to the invention may be any host cell. For specific uses of a compound produced in a host cell according to the invention, the selection of the host cell may be made according to such use. Where e.g. the compound produced in a host cell according to the invention is to be used in food applications, a host cell may be selected from a food-grade organism such as *Saccharomyces cerevisiae*. Specific uses include, but are not limited to, food, (animal) feed, pharmaceutical, agricultural such as crop-protection, and/or personal care applications.

According to an embodiment, the host cell according to the invention is a eukaryotic host cell. Preferably, the eukaryotic cell is a mammalian, insect, plant, fungal, or algal cell. Preferred mammalian cells include e.g. Chinese hamster ovary (CHO) cells, COS cells, 293 cells, PerC6 cells, and hybridomas. Preferred insect cells include e.g. Sf9 and Sf21 cells and derivatives thereof. More preferably, the eukaryotic cell is a fungal cell, i.e. a yeast cell, such as *Candida, Hansenula, Kluyveromyces, Pichia, Saccharomyces, Schizosaccharomyces*, or *Yarrowia strain.* More preferably from *Kluyveromyces lactis*, *S. cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica* and *Pichia pastoris*, or a filamentous fungal cell. Most preferably, the eukaryotic cell is a filamentous fungal cell.

"Filamentous fungi" include all filamentous forms of the subdivision *Eumycota* and *Oomycota* (as defined by Hawksworth *et al.*, In, Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK). The filamentous fungi are characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. Filamentous fungal strains include, but are not limited to, strains of *Acremonium, Agaricus, Aspergillus, Aureobasidium, Chrysosporium, Coprinus, Cryptococcus, Filibasidium, Fusarium, Humicola, Magnaporthe, Mucor, Myceliophthora, Neocallimastix, Neurospora, Paecilomyces, Penicillium, Piromyces,*

Panerochaete, Pleurotus, Schizophyllum, Talaromyces, Thermoascus, Thielavia, Tolypocladium, and *Trichoderma*.

Preferred filamentous fungal cells belong to a species of an *Aspergillus, Chrysosporium, Penicillium, Talaromyces, Fusarium* or *Trichoderma* genus, and most preferably a species of *Aspergillus niger*, *Aspergillus awamori, Aspergillus foetidus, Aspergillus sojae*, *Aspergillus fumigatus, Talaromyces emersonii*, *Aspergillus oryzae*, *Chrysosporium lucknowense, Fusarium oxysporum*, *Trichoderma reesei* or *Penicillium chrysogenum*. A more preferred host cell is *Aspergillus niger*. When the host cell according to the invention is an *Aspergillus niger* host cell, the host cell preferably is CBS 513.88, CBS124.903 or a derivative thereof. Most preferably, the recombinant host cell according to the invention is a host cell as described in the examples herein, preferably a host cell as in examples 4-11.

Several strains of filamentous fungi are readily accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL) *Aspergillus niger* CBS 513.88, CBS124.903, *Aspergillus oryzae* ATCC 20423, IFO 4177, ATCC 1011, CBS205.89, ATCC 9576, ATCC14488-14491, ATCC 11601, ATCC12892, *P. chrysogenum* CBS 455.95, P*enicillium citrinum* ATCC 38065, *Penicillium chrysogenum* P2, *Talaromyces emersonii* CBS 124.902, *Acremonium chrysogenum* ATCC 36225 or ATCC 48272, *Trichoderma reesei* ATCC 26921 or ATCC 56765 or ATCC 26921, *Aspergillus sojae* ATCC11906, *Chrysosporium lucknowense* ATCC44006 and derivatives thereof.

According to another embodiment, the host cell according to the invention is a prokaryotic cell. Preferably, the prokaryotic host cell is bacterial cell. The term “bacterial cell” includes both Gram-negative and Gram-positive microorganisms. Suitable bacteria may be selected from *e.g. Escherichia*, *Anabaena*, *Caulobactert*, *Gluconobacter*, *Rhodobacter*, *Pseudomonas*, *Paracoccus*, *Bacillus*, *Brevibacterium*, *Corynebacterium*, *Rhizobium* (*Sinorhizobium*), *Flavobacterium*, *Klebsiella*, *Enterobacter*, *Lactobacillus*, *Lactococcus, Methylobacterium, Staphylococcus* or *Streptomyces*. Preferably, the bacterial cell is selected from the group consisting of *B. subtilis, B. amyloliquefaciens,*

B. licheniformis, *B. puntis*, *B. megaterium, B. halodurans*, *B. pumilus*, *G. oxydans, Caulobactert crescentus CB 15, Methylobacterium extorquens, Rhodobacter sphaeroides, Pseudomonas zeaxanthinifaciens*, *Paracoccus denitrificans*, *E. coli*, *C. glutamicum*, *Staphylococcus carnosus*, *Streptomyces lividans*, *Sinorhizobium melioti* and *Rhizobium radiobacter*.

The host cell according to the invention can conveniently be used for various purposes, e.g. the production of a compound of interest and for expression cloning such as described in WO1999/032617 and WO2008/138835. These patent applications describe the convenience of expression cloning using filamentous fungal cells as host cells.

The present invention also provides for a method for the production of a recombinant host cell for the production of a compound of interest, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest, comprising adapting a substantially homologous DNA domain with higher gene conversion ratio compared to another substantially homologous DNA domain to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, preferably by providing the substantially homologous DNA domain with higher gene conversion frequency with a targeting DNA domain, wherein said targeting DNA domain comprises a sequence with enhanced integration preference.

Adapting a substantially homologous DNA domain with higher gene conversion ratio compared to another substantially homologous DNA domain to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, may be performed by any means known to the person skilled in the art. Preferably, the adapted substantially homologous DNA domain is partly or completely unique within the genome of the host. As is shown in example 3, an amplicon in *A. niger* is adapted using the glucoamylase promoter sequence, which was previously removed from the genome. This resulted in a substantially homologous DNA domain with a unique targeting sequence. The promoter and terminator parts of a polynucleotide of interest function as DNA sequences suitable

for integration in the adapted substantially homologous DNA domain after transformation as is shown in example 11. All means for performing these experiments for the person skilled in the art are described herein and in Sambrook and Russel, *supra*.

Preferably, the method for the production of a recombinant host cell for the production of a compound of interest further comprises the steps of:

a. transforming the cell obtained with adapted substantially homologous DNA domain with higher gene conversion ratio with a polynucleotide of interest,
b. selecting or screening for cells having at least one copy of the polynucleotide of interest integrated into at least one of the substantially homologous DNA domains,
c. propagating the cells obtained in (b) and selecting or screening for cells having at least one copy of said polynucleotide of interest integrated in an additional copy of a substantially homologous DNA domain.

More preferably, the method for the production of the recombinant host cell for the production of a compound of interest is a method as described in the examples herein.

Methods for transformation of cells are well-known to the person skilled in the art. Any such method may be used for the purpose of the present invention. Transformation may involve a process consisting of protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known per se. Suitable procedures for transformation of *Aspergillus* cells are described in EP 238 023 and Yelton et al., 1984, Proceedings of the National Academy of Sciences USA 81 : 1470-1474. Suitable procedures for transformation of *Aspergillus* and other filamentous fungal host cells using *Agrobacterium tumefaciens* are described in e.g. Nat Biotechnol. 1998 Sep;16(9):839-42. Erratum in: Nat Biotechnol 1998 Nov;16(11):1074. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. de Groot MJ, Bundock P, Hooykaas PJ, Beijersbergen AG. Unilever Research Laboratory Vlaardingen, The Netherlands. A suitable method of transforming *Fusarium* species is described by Malardier et al., 1989, Gene 78: 147156 or in WO 96/00787. Other method can be applied such as a method using biolistic transformation as described in:

Biolistic transformation of the obligate plant pathogenic fungus, Erysiphe graminis f.sp. hordei. Christiansen SK, Knudsen S, Giese H. Curr Genet. 1995 Dec; 29(1):100-2. Yeast may be transformed using the procedures described by Becker and Guarente, In Abelson, J. N. and Simon, M. I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito et al., 1983, Journal of Bacteriology 153: 163; and Hinnen et al., 1978, Proceedings of the National Academy of Sciences USA 75: 1920.

A number of routine techniques are available to the skilled person for determining which of the obtained transformants has an integration of a recombinant DNA molecule in one of its DNA domains.

In a further step a selected transformant is propagated and from its progeny a strain is selected in which at least two of the DNA domains comprise the integrated polynucleotide of interest. This means that strains are selected in which the DNA domain comprising the integrated polynucleotide of interest is multiplied, either through gene conversion with an "empty" DNA domain or through amplification. Such gene conversion and/or amplification events occur spontaneously at low frequency. The exact frequency at which these events occur may depend on a number of variables including the host cell in question and the number, the length and the extent of homology of the DNA domains. These frequencies are however sufficiently high to enable one to screen and select for strains in which these events have occurred using analysis techniques known to the person skilled in the art. Strains in which the DNA domain comprising the integrated polynucleotide of interest is multiplied can e.g. be identified by simply screening for strains with higher production levels of the product encoded by the polynucleotide of interest, or alternatively by analysing their genotype by e.g. the "DNA-flag" test as outlined above.

The method for the production of a recombinant host cell according to the invention may comprise additional steps in which one of the strains in which multiplication of the DNA domain comprising the integrated polynucleotide of interest has occurred is propagated and wherein form its progeny strains are selected in which additional copies of the DNA domains comprise the integrated polynucleotide of interest. These strains may then again be subjected to this procedure until a strain is

obtained in which each of the DNA domains comprises the integrated polynucleotide of interest.

Preferably, step (c) of the method here above is repeated until at least three of the substantially homologous DNA domains have at least one copy of the polynucleotide of interest integrated. More preferably, step (c) is repeated until at least four of the substantially homologous DNA domains have at least one copy of the polynucleotide of interest integrated. Even more preferably, step (c) is repeated until at least five of the substantially homologous DNA domains have at least one copy of the polynucleotide of interest integrated. Even more preferably, step (c) is repeated until at least six of the substantially homologous DNA domains have at least one copy of the polynucleotide of interest integrated. Most preferably, step (c) is repeated until each of the substantially homologous DNA domains has at least one copy of the polynucleotide of interest integrated.

The present invention also provides for a method for the production of a compound of interest, comprising:

a. cultivating a host cell according to the invention under conditions conducive to the production of said compound; and
b. recovering the compound of interest from the cultivation medium.

The present invention also provides for a method for the production of a compound of interest comprising:

a. cultivating a recombinant host cell under conditions conducive to the production of said compound, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest, wherein at least one of the at least two substantially homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, and wherein at least two of the substantially homologous DNA domains have at least one copy of a polynucleotide of interest integrated; and
b. recovering the compound of interest from the cultivation medium.

In the production methods according to the present invention, the host cells are cultivated in a nutrient medium suitable for production of the compound of interest, e.g. polypeptide or metabolite using methods known in the art. Examples of cultivation methods which are not construed to be limitations of the invention are submerged fermentation, surface fermentation on solid state and surface fermentation on liquid substrate. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the coding sequence to be expressed and/or the polypeptide to be isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g., in catalogues of the American Type Culture Collection). If the polypeptide or metabolite is secreted into the nutrient medium, the polypeptide or metabolite can be recovered directly from the medium. If the polypeptide or metabolite is not secreted, it can be recovered from cell lysates.

Polypeptides may be detected using methods known in the art that are specific for the polypeptides. These detection methods may include use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme substrate.

The resulting compound of interest e.g. polypeptide or metabolite may be recovered by the methods known in the art. For example, the polypeptide or metabolite may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation.

Polypeptides may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing), differential solubility (e.g., ammonium sulphate precipitation), SDS-PAGE, or extraction (see, e.g., Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Preferably, the host cell in the methods according to the invention is a selection marker free host cell as described earlier herein.

Preferably, the host cell in the methods according to the invention is a fungal host cell as described earlier herein.

Preferably, the host cell in the methods according to the invention is a filamentous fungal host cell as described earlier herein.

The compound of interest of the present invention can be any biological compound. The biological compound may be any biopolymer or metabolite. The biological compound may be encoded by a single polynucleotide or a series of polynucleotides composing a biosynthetic or metabolic pathway or may be the direct result of the product of a single polynucleotide or products of a series of polynucleotides. The biological compound may be native to the host cell or heterologous.

The term "heterologous biological compound" is defined herein as a biological compound which is not native to the cell; or a native biological compound in which structural modifications have been made to alter the native biological compound.

The term "biopolymer" is defined herein as a chain (or polymer) of identical, similar, or dissimilar subunits (monomers). The biopolymer may be any biopolymer. The biopolymer may for example be, but is not limited to, a nucleic acid, polyamine, polyol, polypeptide (or polyamide), or polysaccharide.

The biopolymer may be a polypeptide. The polypeptide may be any polypeptide having a biological activity of interest. The term "polypeptide" is not meant herein to refer to a specific length of the encoded product and, therefore, encompasses peptides, oligopeptides, and proteins. Polypeptides further include naturally occurring allelic and engineered variations of the above- mentioned polypeptides and hybrid polypeptides. The polypeptide may native or may be heterologous to the host cell. The polypeptide may be a collagen or gelatin, or a variant or hybrid thereof. The polypeptide may be an antibody or parts thereof, an antigen, a clotting factor, an enzyme, a hormone or a hormone variant, a receptor or parts thereof, a regulatory protein, a structural protein, a reporter, or a transport protein, protein involved in secretion process, protein involved in folding process, chaperone, peptide amino acid transporter, glycosylation factor, transcription factor,

synthetic peptide or oligopeptide, intracellular protein. The intracellular protein may be an enzyme such as, a protease, ceramidases, epoxide hydrolase, aminopeptidase, acylases, aldolase, hydroxylase, aminopeptidase, lipase. The polypeptide may be an enzyme secreted extracellularly. Such enzymes may belong to the groups of oxidoreductase, transferase, hydrolase, lyase, isomerase, ligase, catalase, cellulase, chitinase, cutinase, deoxyribonuclease, dextranase, esterase. The enzyme may be a carbohydrase, e.g. cellulases such as endoglucanases, β-glucanases, cellobiohydrolases or β-glucosidases, hemicellulases or pectinolytic enzymes such as xylanases, xylosidases, mannanases, galactanases, galactosidases, pectin methyl esterases, pectin lyases, pectate lyases, endo polygalacturonases, exopolygalacturonases rhamnogalacturonases, arabanases, arabinofuranosidases, arabinoxylan hydrolases, galacturonases, lyases, or amylolytic enzymes; hydrolase, isomerase, or ligase, phosphatases such as phytases, esterases such as lipases, proteolytic enzymes, oxidoreductases such as oxidases, transferases, or isomerases. The enzyme may be a phytase. The enzyme may be an aminopeptidase, asparaginase, amylase, carbohydrase, carboxypeptidase, endo-protease, metallo-protease, serine-protease catalase, chitinase, cutinase, cyclodextrin glycosyltransferase, deoxyribonuclease, esterase, alpha-galactosidase, beta-galactosidase, glucoamylase, alpha-glucosidase, beta-glucosidase, haloperoxidase, protein deaminase, invertase, laccase, lipase, mannosidase, mutanase, oxidase, pectinolytic enzyme, peroxidase, phospholipase, polyphenoloxidase, ribonuclease, transglutaminase, or glucose oxidase, hexose oxidase, monooxygenase.

In the methods of the present invention, a polypeptide can also be a fused or hybrid polypeptide to which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide or fragment thereof. A fused polypeptide is produced by fusing a nucleic acid sequence (or a portion thereof) encoding one polypeptide to a nucleic acid sequence (or a portion thereof) encoding another polypeptide.

Techniques for producing fusion polypeptides are known in the art, and include, ligating the coding sequences encoding the polypeptides so that they are in frame and expression of the fused polypeptide is under control of the same promoter (s) and terminator. The hybrid polypeptides may comprise a combination of partial or complete polypeptide sequences obtained from at least two different polypeptides wherein one or more may be heterologous to the host cell.

The biopolymer may be a polysaccharide. The polysaccharide may be any polysaccharide, including, but not limited to, a mucopolysaccharide(e. g., heparin and hyaluronic acid) and nitrogen-containing polysaccharide (eg., chitin). In a more preferred option, the polysaccharide is hyaluronic acid.

The polynucleotide of interest according to the invention may encode an enzyme involved in the synthesis of a primary or secondary metabolite, such as organic acids, carotenoids, (beta-lactam) antibiotics, and vitamins. Such metabolite may be considered as a biological compound according to the present invention.

The term "metabolite" encompasses both primary and secondary metabolites; the metabolite may be any metabolite. Prefered metabolites are citric acid, gluconic acid and succinic acid.

The metabolite may be encoded by one or more genes, such as in a biosynthetic or metabolic pathway. Primary metabolites are products of primary or general metabolism of a cell, which are concerned with energy metabolism, growth, and structure. Secondary metabolites are products of secondary metabolism (see, for example, R. B. Herbert, The Biosynthesis of Secondary Metabolites, Chapman and Hall, New York, 1981).

The primary metabolite may be, but is not limited to, an amino acid, fatty acid, nucleoside, nucleotide, sugar, triglyceride, or vitamin.

The secondary metabolite may be, but is not limited to, an alkaloid, coumarin, flavonoid, polyketide, quinine, steroid, peptide, or terpene. The secondary metabolite may be an antibiotic, antifeedant, attractant, bacteriocide, fungicide, hormone, insecticide, or rodenticide. Preferred antibiotics are cephalosporins and beta-lactams.

The biological compound may also be the product of a selectable marker. A selectable marker is a product of a polynucleotide of interest which product provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Selectable markers include, but are not limited to, amdS (acetamidase), argB (ornithinecarbamoyltransferase), bar (phosphinothricinacetyltransferase), hygB (hygromycin phosphotransferase), niaD (nitratereductase), pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), trpC (anthranilate synthase), ble (phleomycin resistance protein), as well as equivalents thereof.

According to the invention, the compound of interest in the methods according to the invention is preferably a polypeptide as described herein.

Preferably, the polypeptide in the methods according to the invention is an enzyme as described herein.

According to the invention, the compound of interest in the methods according to the invention is preferably a metabolite.

According to another aspect of the invention, the adapted DNA domain of the at least two substantially homologous DNA domains of the host cell has the lowest frequency of multiplication through gene conversion and / or amplification. Different substantially homologous DNA domains revealed to have different frequencies of multiplication i.e. gene conversion frequencies. The advantage of adapting a targeting DNA domain of a substantially homologous DNA domain with the lowest frequency of multiplication is that a genetically stable host is obtained with a low frequency of recombination in a substantially homologous DNA domain.

The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The specific sequences disclosed herein can be readily used to isolate the complete gene from the respective host cells which in turn can easily be subjected to further sequence analyses thereby identifying sequencing errors.

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a nucleic acid sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a

frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The person skilled in the art is capable of identifying such erroneously identified bases and knows how to correct for such errors.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein enclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. In case of conflict, the present disclosure including definitions will be taken as a guide.

Examples

Strains

WT 1: This *Aspergillus niger* strain is used as a wild-type strain. This strain is deposited at the CBS Institute under the deposit number CBS 513.88.

WT 2: This *A. niger* mutant strain is derived from *A. niger* CBS 513.88 by classical strain improvement by a method as essentially as described in WO98/46772. After NTG- or UV-mutagenesis of spores of WT 1, a selection was performed for improved glucoamylase production of the mutant strains in shake flask. Several improved *A. niger* strains were identified, of which a good one produced a 3-4 fold increased glucoamylase activity levels. This mutant strain is deposited at the CBS Institute under the deposit number CBS 124.903. From shake flask and genetic analyses it was concluded that the *A. niger* CBS 124.903 strain has a 3-4 fold increased glucoamylase production accompanied by an increased (3) *gla*A gene copy number due to amplified *gla*A loci (amplicons).

GBA 300: This *A. niger* strain is a WT 2 strain comprising three modifications of the *gla*A amplicon. Construction of the GBA 300 strain was performed according to the methods described in WO98/46772; in this patent application it is extensively described how to delete three *gla*A specific DNA sequences from an *A. niger* genome containing amplicons, resulting in an *amd*S-negative strain with three truncated *gla*A amplicons. The procedure resulted in a MARKER-GENE FREE Δ*gla*A recombinant *A. niger* WT 2 strain, possessing finally no foreign DNA sequences at all. The three truncated *gla*A amplicons are designated as *Bam*HI truncated amplicon, *Sal*I truncated amplicon and *Bgl*II truncated amplicon.

A. niger shake flask fermentations

A. niger strains were precultured in 20 ml preculture medium as described in the Examples: "Aspergillus niger shake flask fermentations" section of WO99/32617. After overnight growth, 10 ml of this culture was transferred to fermentation medium 1 (FM1) with 7 % glucose as described in WO99/32617. This FM1 contains per liter: 25 g Caseinhydrolysate, 12.5 g Yeast extract, 1 g KH_2PO_4, 2 g K_2SO_4, 0.5 g $MgSO_4.7H_2O$, 0.03 g $ZnCl_2$, 0.02 g $CaCl_2$, 0.01 g $MnSO_4.4H_2O$, 0.3 g $FeSO_4.7H_2O$, 10 ml Pen-Strep (5000 IU/ml Pen-5 mg/ml Strep), adjusted to pH 5.6 with 4 N H_2SO_4. Fermentation is performed in 500 ml flasks with baffle with 100 ml fermentation broth at 34°C and 170 rpm for the number of days indicated.

Fermentation medium 2 (FM2) is used for PLA2 fermentations and contains per liter: 82.5 g Glucose.$1H_2O$, 25 g Maldex 15 (Boom Meppel, Netherlands), 2 g Citric acid, 4.5 g $NaH_2PO_4.1H_2O$, 9 g KH_2PO_4, 15 g $(NH_4)_2SO_4$, 0.02 g $ZnCl_2$, 0.1 g $MnSO_4.1H_2O$, 0.015 g $CuSO_4.5H_2O$, 0.015 g $CoCl_2.6H_2O$, 1 g $MgSO_4.7H_2O$, 0.1 g $CaCl_2.2H_2O$, 0.3 g $FeSO_4.7H_2O$, 30 g MES (2-[N-Morpholino]ethanesulfonic acid), pH=6.

A. niger agar media for toxin measurements

Strains were cultured on the following media to test for the production of ochratoxin A (OTA) and fumonisin B2 (FB2): Czapek yeast autolysate agar (CYA) and yeast extract sucrose agar (YES) as essentially described by Frisvad and Filtenborg, 1989 (Terverticillate Penicillia: chemotaxonomy and mycotoxin production, Mycologia 81:837-861) and Frisvad and Thrane, 1993 (Liquid column chromatography of mycotoxinsm In: Betina (ed): Chromatography of mycotoxins. Techniques and

applications. J. of Chromatography Library 54: 253-372 Elsevier, Amsterdam). Petri dishes were incubated for 7 days in darkness at 24°C, after which agar plugs were taken from single colonies for extraction. Extraction was done with 84% CH3CN-water and placement in an ultasonication bath for 1 hour, after which the solvent was filtrated through a 0.45 µm PTFE filter.

Quantification was done by spiking samples with an internal standard solution to defined amounts of FB2 and OTA. LC-MS/MS was performed as described by Mogensen *et al*. 2010 (Production of fumonisin B2 and B4 by *Aspergillus niger* in raisins and grapes, J of Agricultural and Food Chemistry 58: 954-958) and reviewed in Nielsen *et al*. 2009 (Review of secondary metabolites and mycotoxins from the *Aspergillus niger* group, Anal Bioanal Chem 395: 1225-1242).

Example 1. Construction of *Aspergillus niger* GBA 301 (Δ*gla*A, Δ*pep*A)

This *A. niger* strain is a GBA 300 strain comprising a deletion of the *pep*A gene encoding the major extracellular aspartic protease PepA. The GBA 301 strain is constructed by using the "MARKER-GENE FREE" approach as described in EP 0 635 574. The method described in this patent is used to delete *pep*A specific DNA sequences in the genome of GBA 300, as described by van den Hombergh et al. (van den Hombergh JP, Sollewijn Gelpke MD, van de Vondervoort PJ, Buxton FP, Visser J. (1997) - Disruption of three acid proteases in Aspergillus niger--effects on protease spectrum, intracellular proteolysis, and degradation of target proteins - Eur J Biochem. 247(2): 605-13). The procedure resulted in a MARKER-GENE FREE GBA 301 strain, with the *pep*A gene inactivated in the GBA 300 strain background.

Example 2. Construction of *Aspergillus niger* GBA 302 (Δ*gla*A, Δ*pep*A, Δ*hdf*A)

A gene replacement vector for *hdf*A were designed according to known principles and constructed according to routine cloning procedures. In essence, these vectors comprise approximately 1 – 2 kb flanking regions of the *hdf*A ORF for homologous recombination at the predestined genomic loci. In addition, they contain the *A. nidulans* bi-directional *amd*S selection marker, in-between direct repeats. The general design of deletion vectors was previously described in EP635574B and WO

98/46772 and use of general cloning vector pGBDEL (Figure 1) for constructing deletion vectors was described in WO06/040312.

Vector pDEL-HDFA (Figure 2) comprises approximately 1 kb flanking regions of the *hdf*A ORF for homologous recombination. All nucleotide sequences for *A. niger* genes and their genomic context can be derived for example from NCBI (http://www.ncbi.nlm.nih.gov/) or EMBL (http://www.ebi.ac.uk/embl/). Linear DNA of deletion vector pDEL-HDFA was isolated and used to transform *Aspergillus niger* GBA 301 using a method as earlier described in detail in WO05/095624 to delete the *hdf*A gene. The method applied for gene deletion in all examples herein used linear DNA, which integrates into the genome at the homologous locus of the flanking sequences by a double cross-over, thus substituting the gene to be deleted by the *amd*S gene. Subsequently, the *amd*S marker was removed by plating on fluoro-acetamide media, to select marker-gene-free strains. The general procedure for gene disruption is depicted in Figure 3. Strain GBA 302 was selected as a representative strain with the *hdf*A gene inactivated in the GBA 301 strain background.

Example 3. Construction of *Aspergillus niger* GBA 303 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon)

In this example, the Δ*gla*A amplicons with the highest frequency of gene conversion will be adapted, resulting in a unique locus for targeted integration allowing the use of a smart integration strategy. The adapted amplicon allows the flanking control sequences, such as the glucoamylase promoter and terminator in this example, to be used as targeting regions, as well.

Example 3.1 Identification of amplicon with highest frequency of gene conversion

Strain GBA 301 contains three modified Δ*gla*A loci in the genome. For all three loci, approximately 4.3 kb *gla*A sequences (2 kb *gla*A promoter and entire *gla*A coding sequence) have been deleted (As in Figure 4 – from WO98/46772). Since three different deletion vectors were used, each Δ*gla*A locus is slightly different, features which can be used to visualize each truncated Δ*gla*A locus by a PCR test (Figure 5 – from WO98/46772). This incorporated difference also can be used to follow gene conversion events between the truncated *gla*A amplicons. The three truncated *gla*A

amplicons are designated as *Bam*HI truncated amplicon, *Sal*I truncated amplicon and *Bgl*II truncated amplicon and showing a band with a size of 240, 260 and 300 bp, respectively in the so-called "DNA-flag" test (Figure 5 – from WO98/46772).

From strain GBA 301, multiple phytase producing strains were produced, containing multiple phytase expression cassettes (by cotransforming pGBAAS-1 and pGBTOPFYT-1) all targeted to one of the three truncated *gla*A amplicons of GBA 301 (essentially as described in WO98/46772 – example 1.4). The resulting strains each possess phytase cassettes in one of the three truncated *gla*A amplicons and are called BAM-PHY, SAL-PHY and BGL-PHY, according to the phytase cassette integration site.

Convertants with an increased phytase cassette copy number were identified for the three different strains BAM-PHY, SAL-PHY and BGL-PHY (essentially as described in WO98/46772 – example 1.5). For the convertants with increased phytase copy number, the frequencies and genotypes of specific amplicons of the phytase-containing *gla*A amplicon were determined using the "DNA-flag" test. By analysing thousands of individual progenies of the respective strains by the "DNA-flag" test, gene conversion can be detected by deletion of a specific truncated *gla*A amplicon (for example *Sal*I) paralleled by amplification of another specific truncated *gla*A amplicon (for example *Bam*HI).

The results of PCR-scoring and typing of hundreds of individual progenies for the three different strains can be found in Table 1. In addition, a similar approach was followed for some isolated strains after a single gene conversion event, which contained 2 amplicons of the same type. These results of PCR-scoring hundreds of individual progenies for the three types of strains can be found in Table 2. For the three different strains, the phytase cassettes located in the BamHI amplicon can be multiplied at the highest frequency by gene conversion, both for a first and a second gene conversion event.

Table 1. Strains with increased copy number as a result of gene conversion as determined by the "DNA-flag" test.

Strain	Genotype	Frequency	Genotype	Frequency
BAM-PHY	*Bam*HI $^{2+}$/*Sal*I $^{+}$/*Bgl*II $^{-}$	0.8 %	*Bam*HI $^{2+}$/*Sal*I $^{-}$/*Bgl*II $^{+}$	0.5 %
SAL-PHY	*Sal*I $^{2+}$/*Bam*HI $^{+}$/*Bgl*II $^{-}$	0.2 %	*Sal*I $^{2+}$/*Bam*HI $^{-}$/*Bgl*II $^{+}$	0.2%

BGL-PHY	*Bgl*II $^{2+}$/*Sal*I $^{+}$/*Bam*HI $^{-}$	0 %	*Bgl*II $^{2+}$/*Sal*I $^{-}$/*Bam*HI $^{+}$	0.1 %

Table 2. Strains with increased copy number as a result of a second gene conversion or amplification event as determined by the "DNA-flag" test.

Strain	Genotype	Frequency
*Bam*HI $^{2+}$/*Sal*I $^{+}$/*Bgl*II $^{-}$ or *Bam*HI $^{2+}$/*Sal*I $^{-}$/*Bgl*II $^{+}$	*Bam*HI $^{3+}$	0.2 %
*Sal*I $^{2+}$/*Bam*HI $^{+}$/*Bgl*II $^{-}$ or *Sal*I $^{2+}$/*Bam*HI $^{-}$/*Bgl*II $^{+}$	*Sal*I $^{3+}$	0.1 %
*Bgl*II $^{2+}$/*Sal*I $^{+}$/*Bam*HI $^{-}$ or *Bgl*II $^{2+}$/*Sal*I $^{-}$/*Bam*HI $^{+}$	*Bgl*II $^{3+}$	0 %

Example 3.2 Adaptation of *Bam*HI amplicon, which has the highest frequency of gene conversion

In this part of the example, it is described how the previously removed glucoamylase promoter P*gla*A (upon threefold modification of the *gla*A loci in CBS 124.903 resulting in strain GBA300) is re-introduced at a single locus in the genome but at a different position in the amplicons to be adapted, resulting in a unique and improved locus for targeted integration.

To be able to adapt an amplicon, a gene replacement type of vector for the Δ*gla*A loci was designed according to known principles and constructed according to routine cloning procedures. For this purpose, the vector pGBGLA-65 was constructed (Figure 6). Basically, this vector contains an *amd*S selection marker in between two P_{glaA} fragments and the 3'*gla*A and 3"*gla*A region. The 3'*gla*A and 3"*gla*A region are used for targeting and integration into the identical genomic region of the Δ*gla*A loci. The two P_{glaA} fragments are used for looping out the amdS selection marker upon counterselection (Figure 3). One PglaA fragment is a truncated P_{glaA} promoter fragment (missing the last 600 bp of the P_{glaA} promoter 3' of the *Mlu*I site), which remains present in the genome after amdS counterselection.

To be able to adapt the *Bam*HI amplicon in the genome, vector pDEL-GLA65 (Figure 6) was digested with *Hin*dIII and gel-purified to remove the *E. coli* backbone.

Linear DNA of deletion vector pDEL-GLA65 was used to transform *Aspergillus niger* GBA 302 using a method as earlier described in detail in WO05/095624. Linear DNA of pDEL-GLA65 can integrate into the amplicons at the 3' *gla*A and 3'' *gla*A sequences by a double cross-over. A number of colonies were purified and analysed via PCR for presence of the three *gla*A-loci and integration of the cassette at the *Bam*HI amplicon (target PCR). The targeting frequency for the *Bam*HI amplicon was estimated at 20 - 30 %. In total 7 strains were selected for fluoracetamide counterselection. The fluoracetamide resistant colonies were transferred to PDA and tested by "flag"-PCR. A strain was selected as a representative strain with an adapted *Bam*HI locus, with the *amd*S marker correctly deleted via recombination over the P*gla*A repeats of the integrated pDEL-GLA65 cassette (figure 7). The selected strain was designated GBA 303, which represents the GBA 300 strain background with all 3 Δ*gla*A loci, with the *pep*A and the *hdf*A gene inactivated and the *Bam*HI amplicon adapted. Since the entire *gla*A amplicons are large in size (>80 kb) and although the *Bam*HI amplicon has been adapted (2-6 kb), still fragments with rather substantial homologous DNA domains remain present in the genome.

The sequence of the genomic region comprising a 3'- fragment and 3''-fragment of a Δ*gla*A amplicon (Figure 7) can be found under SEQ ID NO: 1. The sequence of the genomic region comprising a 3'-fragment, a truncated P_{glaA} promoter fragment and a 3''- fragment of the adapted BamHI amplicon can be found under SEQ ID NO: 2.

Example 3.3 Construction of novel integration vectors

In EPA 0635574A1, WO98/46772 and WO99/32617 the integration and expression vectors pGBAAS-1 (pGBGLA-50 in EP0635574) and pGBTOP-8, used in targeted strain construction, and especially targeted strain construction using a single cross-over have been described. Adjustment of these integration vectors is necessary to allow specific targeting to the adapted *Bam*HI amplicon of GBA 302 type of strains.

Of the pGBAAS-1 vector, containing the *amd*S marker and used in co-transformation, two different variants have been constructed for targeted integration into the adapted *Bam*HI amplicon. A vector called pGBAAS-3 (Figure 8) was constructed, which has the *amd*S marker gene expressed under control of the glucoamylase promoter P_{glaA}. The P_{glaA} promoter has a function in driving the

expression of the amdS marker gene. In addition, the P_{glaA} promoter and the 3' *gla*A fragment function as targeting sequences, specifically for the adapted *Bam*HI Δ*gla*A amplicon (Figure 21). A second vector, called pGBAAS-4 (Figure 9), was constructed, which contains the *amd*S marker cassette of pGBAAS-1 (*amd*S under control of the *gpd*A promoter). Here, a truncated P_{glaA} promoter fragment and the 3' *gla*A fragment have a function in targeting to the adapted *Bam*HI Δ*gla*A amplicon (Figure 21).

For the pGBTOP type of vectors, two novel integration and expression vectors have been constructed to allow targeting to the adapted *Bam*HI Δ*gla*A amplicon. The pGBTOP-11 (Figure 10) and pGBTOP-12 (Figure 11) vectors basically are a pGBTOP-8 vector with the 3" *gla*A fragment removed and with additional modification of restriction sites for *E. coli* vector removal, respectively. Here, the P_{glaA} promoter and the 3' *gla*A terminator fragment have a double function: besides their role as promoter and terminator for expressing a gene of interest also a role in targeting to the adapted *Bam*HI Δ*gla*A amplicon (Figure 21).

Example 4. Construction of *Aspergillus niger* GBA 304 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII)

Gene-flanking regions of the *amy*BII ORF, encoding alpha-amylase, were cloned essentially as described in Example 2, resulting in vector pDEL-AMYBII (Figure 12). Linear DNA of deletion vector pDEL-AMYBII was isolated and used to transform *Aspergillus niger* GBA 303 using a method as earlier described in WO05/095624 to delete the *amy*BII gene. A transformant was selected with the *amy*BII ORF removed upon integration of pDEL-AMYBII in the genome at the homologous *amy*BII locus by a double cross-over event. Subsequently, the *amd*S marker was removed by plating on fluoro-acetamide media, to select marker-gene-free strains. Strain GBA 304 was selected as a representative strain with the *amy*BII gene inactivated in the GBA 303 strain background.

Example 5. Construction of *Aspergillus niger* GBA 305 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII, Δ*amy*BI)

Gene-flanking regions of the *amy*BI ORF, encoding alpha-amylase, were cloned essentially as described in Example 2, resulting in vector pDEL-AMYBI (a representative picture for the layout of pDEL-AMYBI can be found in Figure 12). Linear

DNA of deletion vector pDEL-AMYBI was isolated and used to transform *Aspergillus niger* GBA 304 using a method as earlier described in WO05/095624 to delete the *amy*BI gene. A transformant was selected with the *amy*BI ORF removed upon integration of pDEL-AMYBI in the genome at the homologous *amy*BI locus by a double cross-over event. Subsequently, the *amd*S marker was removed by plating on fluoro-acetamide media, to select marker-gene-free strains. Strain GBA 305 was selected as a representative strain with the *amy*BI gene inactivated in the GBA 304 strain background.

Example 6. Construction of *Aspergillus niger* GBA 306 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII, Δ*amy*BI, Δ*amy*A)

Gene-flanking regions of the *amy*A ORF, encoding acid stable alpha-amylase, were cloned essentially as described in Example 2, resulting in vector pDEL-AMYA (a representative picture for the layout of pDEL-AMYA can be found in Figure 12). Linear DNA of deletion vector pDEL-AMYA was isolated and used to transform *Aspergillus niger* GBA 305 using a method as earlier described in WO05/095624 to delete the *amy*A gene. A transformant was selected with the *amy*A ORF removed upon integration of pDEL-AMYA in the genome at the homologous *amy*A locus by a double cross-over event. Subsequently, the *amd*S marker was removed by plating on fluoro-acetamide media, to select marker-gene-free strains. Strain GBA 306 was selected as a representative strain with the *amy*A gene inactivated in the GBA 305 strain background.

Example 7. Construction of *Aspergillus niger* GBA 307 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII, Δ*amy*BI, Δ*amy*A, Δ*oah*A)

This *A. niger* oxalate deficient strain can be obtained by deletion of the *oah*A gene, encoding oxaloacetate hydrolase, which is described in detail in EP1157100 and US6,936,438. Strain GBA 307 was selected as a representative strain with the *oah*A gene inactivated in the GBA 306 strain background.

Alternatively, a mutant strain can be derived from *A. niger* GBA 306 by classical strain improvement as described in WO04/070022 and EP1590444. In these documents, it is extensively described how to screen for an oxalate deficient *A. niger* strain and were isolated according to the methods of Examples 1 and 2 of EP1590444.

Example 8. Construction of *Aspergillus niger* GBA 308 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII, Δ*amy*BI, Δ*amy*A, Δ*oah*A, Δ*fum*B)

In the genome sequence of CBS 513.88, a putative fumonisin gene cluster was identified on the basis of homology to a gene cluster in *Gibberella moniliformis*, encoding the mycotoxin fumonisin (Pel et al., "Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88". Nat Biotechnol. 2007 Feb; 25 (2):221-231).

Gene ID	Description
An01g08620	Strong similarity to fatty acid omega-hydrolase (P450foxy) CYP505 - *Fusarium oxysporum*
An01g06830	Similarity to 3-ketosphinganine reductase Tsc10 – *Saccharomyces cerevisiae*
An01g06840	Strong similarity to acid-CoA ligase Fat2 - *Saccharomyces cerevisiae*
An01g06850	Strong similarity to 4-hydroxybutyrate dehydrogenase – *Alcaligenes eutropus*
An01g06860	Strong similarity to hypothetical protein Fum9p – *Gibberella moniliformis*
An01g06870	Strong similarity to hypothetical protein Fum8p – *Gibberella moniliformis*
An01g06880	Similarity to dihydroflavonol 4-reductase BAA12723.1 – *Rosa hybrid cultivar*
An01g06890	Similarity to peptide synthase pesA – *Metarhizium anisopliae*
An01g06900	Weak similarity to transcription regulator of maltose utilization amyR – *Aspergillus oryzae*
An01g06910	Strong similarity to cytochrome P450 CYP94A5 – *Nicotiane tabacum*
An01g06920	Strong similarity to multidrug resistance protein ABCC2 – *Homo sapiens*
An01g06930 - *fum*B	Strong similarity to polyketide synthase FUM5 - *Gibberella moniliformis*
An01g06940	Strong similarity to hypothetical transmembrane transport protein SCC30.17c – *Streptomyces coelicolor*
An01g06950	Strong similarity to polyketide synthase FUM5 - *Gibberella moniliformis*

The gene An01g06930 encodes a polyketide synthase (PKS) possibly involved in fumonisin production. Gene-flanking regions, with on the promoter side also a small part of the coding sequence of the putative *fum*B gene (An01g06930), were cloned as essentially described in Example 2 and more detailed in WO06/040312, in vector pGBDEL (Figure 1), resulting in vector pGBDEL-FUM3 (Figure 13). Linear DNA of deletion vector pGBDEL-FUM3 was isolated and used to transform *Aspergillus niger* GBA 307 as earlier described in WO06/040312 to delete the *fum*B gene. A

transformant was selected with the *fum*B ORF removed upon integration of pGBDEL-FUM3 in the genome at the homologous An01g06930 locus by a double cross-over event, resulting in the removal of approximately 8 kb of genome sequence. Subsequently, the *amd*S marker was removed by plating on fluoro-acetamide media, to select marker-gene-free strains. Strain GBA 308 was selected as a representative strain with the *fum*B gene inactivated in the GBA 307 strain background.

Strain WT1 and GBA 308 were grown on CYA and YES agar media and fumonisin production was measured. The results of the FB2 measurements are indicated in the Table below.

Strain	FB2 on CYA (ng/cm2)	FB2 on YES (ng/cm2)
WT 1	772	218
GBA 308	0	0

From these results, it is clear that disruption of the PKS-encoding *fum*B gene results in strains with impaired fumonisin production. A fumonisin-negative strain background has a clear benefit for commercial protein production.

In addition it was shown that strains with disruption of a large gene cluster comprising genes An01g06820 untill An01g06930 (having approximately 38 kb of genome sequence removed) shows an identical phenotype, *i.e.* being negative in fumonisin production in the test as detailed above (data not shown).

Example 9. Construction of *Aspergillus niger* GBA 309 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII, Δ*amy*BI, Δ*amy*A, Δ*oah*A, Δ*fum*B, Δ*och*)

In the genome sequence of CBS 513.88, a putative ochratoxin gene cluster was identified on the basis of a PKS fragment of *A. ochraceus* involved in ochratoxin production (Pel et al., “Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88”. Nat Biotechnol. 2007 Feb; 25 (2):221-231).

Gene ID	Description
An15g07860	Strong similarity to hypothetical short chain dehydrogenase SPCC736.13 – *Schizosaccharomyces pombe*
An15g07870	Strong similarity to alcohol dehydrogenase adhT – *Bacillus stereothermophilus*

An15g07880	Strong similarity to hypothetical hydrolase A – *Amycolatopsis orientalis*
An15g07890	Similarity to protein c-fos – Xenopus laevis
An15g07900	Strong similarity to cytochrome P450 – *Myrothecium roridum*
An15g07910 - *och*B	Strong similarity to cyclic peptide AM-toxin synthase AMT – *Alternaria alternaria*
An15g07920 - *och*A;	Strong similarity to PKS of *A. ochraceus* fragment involved in ochratoxin biosynthesis
An15g07930	Strong similarity to nitric oxide synthase – *Manduca sexta* [truncated ORF]

This PKS-like gene is annotated as An15g07920 (called *och*A). Additionally, the gene An15g07910 (called *och*B) could represent a potential peptide synthase involved in ochratoxin production. Disruption of the peptide synthase and/or PKS-encoding genes could lead to strains impaired in ochratoxin production.

To be able to disrupt the full ochratoxin gene cluster using targeting sequences for double cross-over, a gene sequence fragment of the An15g07860 ORF on one side and a gene sequence fragment of the An15g07930 ORF on the other side were cloned as essentially described in Example 2 and more detailed in WO06/040312, resulting in vector pGBDEL-OCH2 (Figure 14). Linear DNA of deletion vector pGBDEL-OCH2 was isolated and used to transform *Aspergillus niger* GBA 308 as earlier described in WO06/040312 to delete the putative ochratoxin gene cluster. A transformant was selected with the *och* ORF's removed upon integration of pGBDEL-OCH2 in the genome at the homologous loci by a double cross-over event. Subsequently, the *amd*S marker was removed by plating on fluoro-acetamide media, to select marker-gene-free strains. Strain GBA 309 was selected as a representative strain with the ochratoxin genomic gene cluster (between gene An15g07860 and An15g07930; genotypically indicated as Δ*och*) removed in the GBA 308 strain background.

Strain WT1 and GBA 309 were grown on CYA and YES agar media and ochrotoxin A production was measured. The results of the OTA measurements are indicated in the Table below.

Strain	OTA on CYA (ng/cm2)	OTA on YES (ng/cm2)
WT 1	0	178
GBA 309	0	0

From these results, it is clear that disruption of the ochrotoxin gene cluster comprising a putative peptide synthase and PKS involved in OTA biosynthesis results in strains with impaired ochrotoxin A production. A ochratoxin A-negative and a possible combination with a fumonisin-negative strain background has a clear benefit for commercial protein production.

Example 10. Construction of *Aspergillus niger* GBA 310 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII, Δ*amy*BI, Δ*amy*A, Δ*oah*A, Δ*fum*B, Δ*och*, Δ*prt*T)

Gene-flanking regions of the *prt*T ORF, encoding a protease regulator, were cloned as essentially described in Example 2 and more detailed in WO06/040312, resulting in vector pGBDEL-PRT2 (Figure 15). Linear DNA of deletion vector pGBDEL-PRT2 was isolated and used to transform *Aspergillus niger* GBA 309 as earlier described in WO06/040312 to delete the *prt*T gene. A transformant was selected with the *prt*T ORF removed upon integration of pGBDEL-PRT2 in the genome at the homologous *prt*T locus by a double cross-over event. Subsequently, the *amd*S marker was removed by plating on fluoro-acetamide media, to select marker-gene-free strains. Strain GBA 310 was selected as a representative strain with the *prt*T gene inactivated in the GBA 309 strain background.

Example 11. Construction of *Aspergillus niger* GBA 311 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII, Δ*amy*BI, Δ*amy*A, Δ*oah*A, Δ*fum*B, Δ*och*, Δ*prt*T, *sec*61 S376W mutation)

In this example, the introduction of a modified *sec*61 gene, called *sec*61* encoding a Sec61* protein in which Serine 376 is replaced by tryptophan, is described. To do so, gene-flankings and s*ec*61* coding sequence, encoding a modified Sec61 translocation channel, were cloned as essentially described in detail in WO2005123763, resulting in vector pGBDEL-SEC61* (Figure 16). Linear DNA of deletion vector pGBDEL-SEC61* was isolated and used to transform *Aspergillus niger* GBA 310 as earlier described in WO2005123763 to modify the *Sec*61 gene. A transformant was selected with a modified *Sec*61* ORF upon integration of pGBDEL-SEC61* in the genome at the homologous *Sec*61 locus by a double cross-over event.

Subsequently, the *amd*S marker was removed by plating on fluoro-acetamide media, to select marker-gene-free strains. Strain GBA 311 was selected as a representative strain with the *Sec*61 gene modified in the GBA 310 strain background.

Example 12. Improved enzyme production and production strain building using adapted *Aspergillus niger* strains and plasmids

Porcine phospholipase A2 (PLA2) protein (Roberts I.N., Jeenes D.J., MacKenzie D.A., Wilkinson A.P., Sumner I.G. and Archer D.B. (1992) "Heterologous gene expression in *Aspergillus niger*: a glucoamylase-porcine pancreatic phospholipase A_2 fusion protein is secreted and processed to yield mature enzyme" Gene 122: 155-161) was selected as a model protein for enzyme expression in the *A. niger* CBS 513.88 & GBA- strain lineage. The fragment for overexpression of PLA2 was made as a fusion of proPLA2 with a native glucoamylase A gene of *A. niger* and was prepared in principle as described by Roberts *et al*. (1992). This *gla*A-*pla*2 fusion gene was cloned into pGBTOP-8 and pGBTOP-12, resulting in pGBTOPPLA-2a (Figure 17) and pGBTOPPLA-2b (Figure 18), respectively.

Additionally, alpha-amylase of *A. niger* was used as a second model protein for enzyme expression in the *A. niger* CBS 513.88 & GBA- strain lineage. The *amy*B coding sequence used for overexpression comprised a single-codon and codon-pair optimized coding sequence for the alpha-amylase encoding *amy*B gene (as described in detail in WO2008/000632). The translational initiation sequence of the glucoamylase glaA promoter has been modified into 5'-CACCGTCAAA ATG-3' in the *amyB* expression constructs generated (as also detailed in WO2006/077258). In addition, an optimal translational termination sequence was used, and therefore the wild-type 5'-TGA-3' translational termination sequence was replaced by 5'-TAAA-3' (as detailed in WO2006/077258) in all expression constructs. An *Eco*RI - *Sna*BI fragment, comprising an optimized amyB cDNA sequence was synthesized completely, subcloned, sequence verified by sequence analysis. The *amy*B fragment was cloned into pGBTOP-8 and pGBTOP-12, resulting in pGBTOPFUA-2 (Figure 19) and pGBTOPFUA-3 (Figure 20), respectively.

Example 12.1 Improved targeting of adapted expression vectors to Δ*gla*A loci in *A. niger* strains

The constructed integration vectors (Example 3.3 and 12) can be divided into two types: The integration vectors that can be used to target to the non-adapted Δ*gla*A loci to the 3' *gla*A and 3"*gla*A sequences, and which are vectors pGBTOPPLA-2a, pGBTOPFUA-2 and pGBAAS-1. Additionally, novel integration vectors have been constructed that can be used to target specifically to the adapted *Bam*HI Δ*gla*A amplicon (an adapted Δ*gla*A locus) to a truncated P_{glaA} promoter fragment and the 3' *gla*A fragment. The vectors that can be used to target to the adapted *Bam*HI amplicon are pGBTOPPLA-2b, pGBTOPFUA-3, pGBAAS-3 and pGBAAS-4.

The two PLA2 expression vectors (pGBTOPPLA-2a and pGBTOPPLA-2b) were introduced by co-transformation with pGBBAAS-1 and pGBAAS-4 in *A. niger* GBA 301 (Δ*gla*A, Δ*pep*A), GBA 302 (Δ*gla*A, Δ*pep*A, Δ*hdf*A) and GBA 303 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon) according the scheme below (Table 3). Transformation and subsequent selection of transformants was carried out as described in WO98/46772 and WO99/32617. In principle, linear DNA of all four vectors was isolated and used to co-transform the three *A. niger* strains. Integration of both linearized plasmids occurs via a single cross-over event at one of the Δ*gla*A loci. Transformants were selected on acetamide media and colony purified according standard procedures. Hundred transformants possessing both the *amd*S marker gene and the PLA2 expression cassette were identified by PCR using *amd*S and PLA2-specific primers for all three strains. On the identified co-transformants, targeting PCR tests were performed (as detailed in Example 1.4 of WO98/46772) to determine the percentage of strains with either the *amd*S and / or PLA2 cassettes integrated in a Δ*gla*A locus.

Additionally, *amd*S marker-gene free strains were selected of identified Δ*gla*A targeted co-transformants by plating on fluoro-acetamide media. A PCR-based DNA 'flag-test' was performed to detect a potential loss of an amplicon in the *amd*S-free progeny. As explained in Example 1.4.f of WO98/46772, this is an indication for a simultaneous loss of the *amd*S marker and the entire *gla*A amplicon through recombination and as such an indication the location of the original cassette(s). This test was used to determine the targeting frequency to the *Bam*HI amplicon in all three strains. Data are shown in Table 3.

Table 3. Targeting frequencies of positive co-transformants, containing a PLA2 expression cassette and an *amd*S marker-gene.

Plasmids	Strain	Targeting to Δ*gla*A loci * (*Bam*HI, *Sal*I, *Bgl*II)	Targeting to *Bam*HI Δ*gla*A locus **
pGBTOPPLA-2 pGBBAAS-1	GBA 301	15%	4 %
pGBTOPPLA-2 pGBBAAS-1	GBA 302	87 %	25 %
pGBTOPPLA-3 pGBBAAS-4	GBA 303	91 %	88 %

* As identified by targeting PCR test

** As identified by DNA flag-PCR test of marker-gene free progenies

As can be concluded from Table 3, the use of adapted expression vectors in combination with a adapted targeting DNA domain (the adapted *Bam*HI amplicon) gives a clear advantage in targeted strain building, resulting in a more focussed and efficient strain construction.

Example 12.2 Improved co-transformation using adapted constructs and *A. niger* strains

The FUA expression vector pGBTOPFUA-3 was introduced by co-transformation with pGBAAS-3 and pGBAAS-4 in *A. niger* GBA 307 (Δ*gla*A, Δ*pep*A, Δ*hdf*A, adapted *Bam*HI amplicon, Δ*amy*BII, Δ*amy*BI, Δ*amy*A, Δ*oah*A), with pGBTOPFUA-2 and pGBAAS as reference set, according the scheme below (Table 4). Transformation and subsequent selection of transformants was carried out as described in WO98/46772 and WO99/32617. In principle, linear DNA of the five vectors was isolated and used to co-transform the *A. niger* GBA 307 strain. Integration of both linearized plasmids occurs via a single cross-over event at the adapted *Bam*HI Δ*gla*A amplicon or Δ*gla*A amplicons, respectively. Transformants were selected on acetamide media and colony purified according standard procedures. The co-transformation percentages (principally as detailed in Example 1.4 of WO98/46772) were determined for 50 transformants of each of the plasmid combinations by PCR using *amd*S and *amy*B-specific (codon-pair optimized *amy*B coding sequence) primers. Data are shown in Table 4.

Table 4. Co-transformation frequencies of a FUA expression cassette and an *amd*S marker-gene cassette.

Co-transforming plasmids	Strain	Co-transformation % *
pGBTOPFUA-2 & pGBAAS-1	GBA 307	42 %
pGBTOPFUA-3 & pGBAAS-4	GBA 307	48 %
pGBTOPFUA-3 & pGBAAS-3	GBA 307	94 %

* As identified by PCR test

With the previously known pGBAAS-1 and its variant for improved targeting pGBAAS-4, typical co-tranformation frequencies are obtained (normally between 25 – 75 %). Surprisingly, as can be concluded from Table 4, the use of the pGBBAAS-3 variant (amdS gene under control of *gla*A promoter) in combination with an adapted targeting construct (such as pGBTOPFUA-3) gives a very high co-transformation frequency. The use of pGBAAS-3, increases co-transformation frequencies dramatically, avoiding the need for selection of co-transformants before further strain manipulations in targeted strain building, and thus resulting in a more efficient strain construction.

Shake-flask analysis in FM1 of these transformants indicated a similar FUA expression level per gene copy for the different constructs expressed in GBA 307.

Example 12.3 Improved strain building in adapted *A. niger* strains

One way of constructing recombinant industrial *A. niger* production strains is to integrate expression cassettes in one of the Δ*gla*A loci and subsequently multiplying the Δ*gla*A locus with integrated expression cassettes through gene conversion or amplification.

Expression cassettes can be improved through several methods as detailed for example in WO 2005/100573, WO2006/077258, WO 2006/092396 or WO2008/000632. The use of gene conversion or amplification in strain building has been described in more detail in for example WO1998/46772. In the previous examples it has been shown how co-transformation and / or targeting to a specific locus can be improved by strain and plasmid adaptation. In this example, the use of gene conversion or amplification in strain construction using novel GBA types of strains will be demonstrated.

The *A. niger* GBA 311 (Example 11) strain was co-transformed with 2 combinations of expression vectors (*i.e.* pGBTOPPLA-2a & pGBBAAS-1 mix; pGBTOPPLA-2b & pGBBAAS-3 mix). Transformation and subsequent selection of transformants was carried out as described in WO98/46772 and WO99/32617. In principle, linear DNA of all vectors used was isolated and used to co-transform the GBA 311 *A. niger* strain. Integration of both linearized plasmids occurs via a single cross-over event at a Δ*gla*A locus (See example 12.1 for details in targeting sequences). Transformants were selected on acetamide media and colony purified according standard procedures. Hundred transformants possessing both the *amd*S marker gene and the PLA2 expression cassette were identified by PCR using *amd*S and PLA2-specific primers. On the identified co-transformants, targeting PCR tests were performed (as detailed in Example 1.4 of WO98/46772) to determine the percentage of strains with either the *amd*S and / or PLA2 cassettes integrated in a Δ*gla*A locus. In addition, colonies were diagnosed for PLA2-copy number using PCR. Copy numbers for suspected high-copy strains were determined more accurately using Southern analysis (WO98/46772). For the mix of pGBTOPPLA-2 & pGBBAAS-1, additional transformants were sceened and purified to isolate a transformant with a higher copy number of PLA2 expression cassettes. After screening 600 individual transformants, a 6-copy *Bgl*II-targeted transformant was isolated. Resulting data are shown in Table 5.

Table 5. Co-transformation frequencies and copy number determination of PLA-2 transformants in the GBA 311 strain background.

Plasmids	Strain	# of strains tested	Targeting to Δ*gla*A locus *	Highest copy number of pla expression cassettes
pGBTOPPLA-2 pGBBAAS-1	GBA 311	100	11% (*Sal*I, *Bgl*II)	3 copies (6 copies after screening 600 colonies)
pGBTOPPLA-3 pGBBAAS-3	GBA 311	100	93 % (*Bam*HI)	7 copies

* As identified by targeting PCR test

** As identified by Southern analysis (WO98/46772)

For the two high-copy strains, i.e. the 6-copy *Bgl*II-targeted strain and the 7-copy *Bam*HI-targeted strain in the GBA300 background, convertants with an increased

PLA2-copy number were isolated (for experimental details see Example 3.1 herein and WO98/46772). Upon screening of 500 individual progenies of both strains, 2 strains with multiplied PLA-2 copy numbers as a result of gene conversion were identified for the *Bam*HI-targeted strain and none for the *Bgl*II-targeted strain. The two convertants contained 13 and 14 PLA-2 copies, respectively. This example and data clearly indicate that adaptation of a genomic targeting DNA domain (that for example has enhanced gene conversion or amplification frequencies) has great impact on strain construction by facilitating and speeding up the process of defined and controlled construction of a multi copy (production) strain. Shake-flask analysis in FM2 of these transformants indicated a similar PLA2 expression level per gene copy for the different constructs expressed in GBA 311.

Applicant's or agent's file reference number **26710-WO-PCT**	International application No.

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description first mentioned on page 2, line 13.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet [X]

Name of depositary institution

CENTRAAL BUREAU VOOR SCHIMMELCULTURES

Address of depositary institution *(including postal code and country)*
Uppsalalaan 8
P.O. Box 85167
NL-3508 AD Utrecht
The Netherlands

Date of deposit 01 July 2009	Accession Number CBS 124.903

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet [ ]

We inform you that the availability of the microorganism identified above, referred to Rule 13bis PCT, shall be effected only by issue of a sample to an expert nominated by the requester until the publication of the mention of grant of the national patent or, where applicable, for twenty years from the date of filing if the application has been refused, withdrawn or deemed to be withdrawn.

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only

[X] This sheet was received with the international application

Authorized officer Pasche, Constantinus

For International Bureau use only

[ ] This sheet was received by the International Bureau on:

Authorized officer

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number **26710-WO-PCT**	International application No.

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description first mentioned on page 23, line 23.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet [X]

Name of depositary institution

CENTRAAL BUREAU VOOR SCHIMMELCULTURES

Address of depositary institution *(including postal code and country)*
Uppsalalaan 8
P.O. Box 85167
NL-3508 AD Utrecht
The Netherlands

Date of deposit 10 August 1988	Accession Number CBS 513.88

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet [ ]

We inform you that the availability of the microorganism identified above, referred to Rule 13bis PCT, shall be effected only by issue of a sample to an expert nominated by the requester until the publication of the mention of grant of the national patent or, where applicable, for twenty years from the date of filing if the application has been refused, withdrawn or deemed to be withdrawn.

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only

[X] This sheet was received with the international application

Authorized officer Pasche, Constantinus

For International Bureau use only

[ ] This sheet was received by the International Bureau on:

Authorized officer

Form PCT/RO/134 (July 1992)

CLAIMS

1. A recombinant host cell for the production of a compound of interest, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest wherein at least one of the at least two substantially homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, and wherein the substantially homologous DNA domain where the adapted substantially homologous DNA domain originates from has a gene conversion frequency that is at least 10% higher than one of the other of the at least two substantially homologous DNA domains.

2. The host cell according to claim 1, wherein the adapted substantially homologous DNA domain is distinguished from the other versions of the substantially homologous DNA domains by means of a unique sequence tag.

3. The host cell according to any of the preceding claims, wherein the adapted substantially homologous DNA domain comprises a targeting DNA domain, wherein said targeting DNA domain comprises a sequence with enhanced integration preference.

4. The host cell according to any one of the preceding claims, wherein at least two of the substantially homologous DNA domains each have at least one copy of a polynucleotide of interest integrated.

5. The host cell according to any one of the preceding claims, further comprising increased efficiency of targeted integration of a polynucleotide to a pre-determined site into the genome of said host cell by steering an integration pathway towards HR.

6. The host cell according to any one of the preceding claims, wherein the at least two substantially homologous DNA domains are loci of glaA or are loci of amyBI, or amyBII).

7. The host cell according to any one of the preceding claims, wherein the polynucleotide of interest is a codon optimized polynucleotide.

8. The host cell according to any one of the preceding claims, further comprising a polynucleotide selected from the group of: glaA, amyA, amyBI, amyBII, oahA, toxin associated polynucleotide and prtT, said polynucleotide comprising a modification, wherein the host cell is deficient in the product encoded by the polynucleotide comprising the modification, compared to the parent cell it originates from when cultivated under comparable conditions.

9. The host cell according to claim 8, further comprising a deficiency in pepA.

10. The host cell according to claim 8 or 9, further comprising a modification in SEC61.

11. The host cell according to any one of the preceding claims, wherein the host cell is a selection marker free host cell.

12. The host cell according to any one of the preceding claims wherein the host cell is a fungal host cell

13. The host cell according to claim 12, wherein the host cell is a filamentous fungal host cell.

14. Method for the production of a recombinant host cell for the production of a compound of interest, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest, comprising adapting a substantially homologous DNA domain with higher gene conversion frequency compared to another

substantially homologous DNA domain to have enhanced integration preference for the polynucleotide of interest compared to the substantially homologous DNA domain it originates from, comprising: providing the substantially homologous DNA domain with higher gene conversion frequency with a targeting DNA domain, wherein said targeting DNA domain comprises a sequence with enhanced integration preference.

15. The method according to claim 14, further comprising:
    a. transforming the cell obtained in claim 14 with a polynucleotide of interest,
    b. selecting or screening for cells having at least one copy of the polynucleotide of interest integrated into at least one of the substantially homologous DNA domains,
    c. propagating the cells obtained in (b) and selecting or screening for cells having at least one copy of said polynucleotide of interest integrated in an additional copy of a substantially homologous DNA domain.

16. The method according to claim 15, further comprising the step of repeating step (c) until each copy of the substantially homologous DNA domains has at least one copy of the polynucleotide of interest integrated.

17. Method for the production of a compound of interest, comprising:
    a. cultivating a host cell according to any on of claims 1-13 under conditions conducive to the production of said compound; and
    b. recovering the compound of interest from the cultivation medium.

18. Method for the production of a compound of interest comprising
    a. cultivating a recombinant host cell under conditions conducive to the production of said compound, said host cell comprising at least two substantially homologous DNA domains suitable for integration of one or more copies of a polynucleotide of interest, wherein at least one of the at least two substantially homologous DNA domains is adapted to have enhanced integration preference for the polynucleotide of interest

compared to the substantially homologous DNA domain it originates from, and wherein at least two of the substantially homologous DNA domains have at least one copy of a polynucleotide of interest integrated; and

b. recovering the compound of interest from the cultivation medium.

19. The method according to claim 17 or 18, wherein the host cell is a selection marker free host cell.

20. The method according to any one of claims 17 to 19, wherein the host cell is a fungal host cell.

21. The method according to claim 20, wherein the host cell is a filamentous fungal host cell.

22. The method according to any one of claims 17 to 21, wherein the compound of interest is a polypeptide.

23. The method according to claim 22, wherein the polypeptide is an enzyme.

24. The method according to any one of claims 17 to 21, wherein the compound of interest is a metabolite.

25. An isolated polynucleotide comprising the polynucleotide according to SEQ ID NO: 2, or a functional fragment thereof.

26. An isolated polynucleotide comprising the polynucleotide according to SEQ ID NO: 3, or a functional fragment thereof.

27. A vector comprising the polynucleotide according to claim 25 or 26.

28. A host cell comprising the polynucleotide according to claim 25 or 26, or comprising the vector according to claim 27.

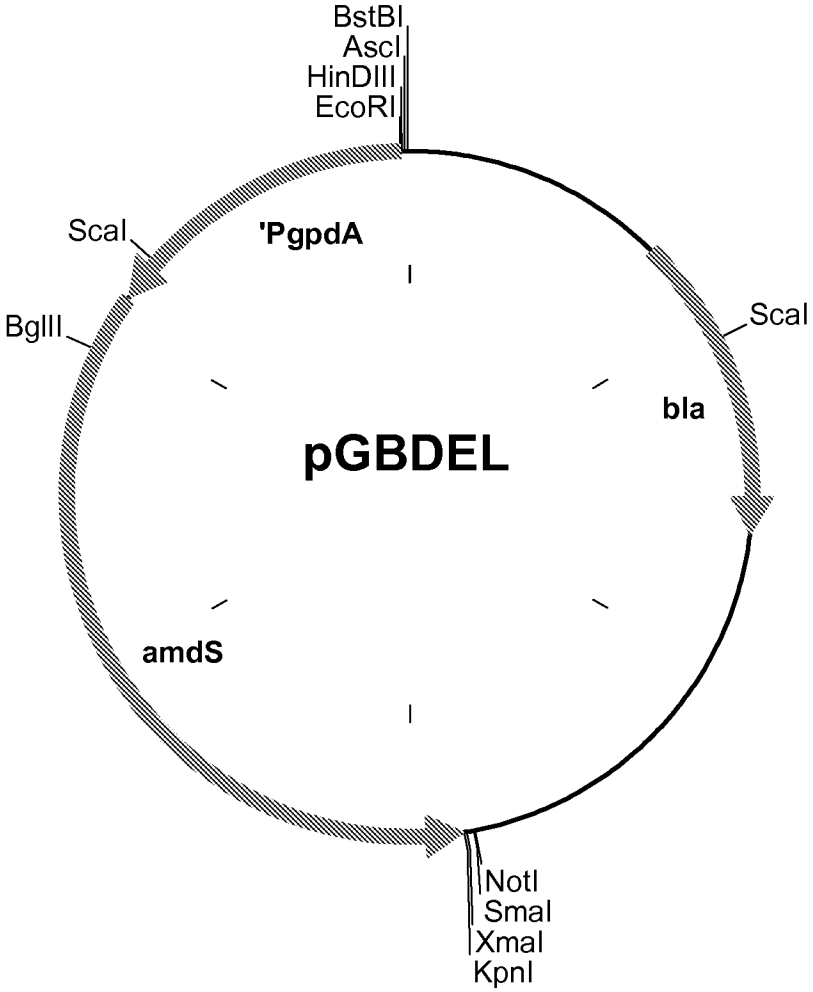

BstBI
AscI
HinDIII
EcoRI
ScaI
'PgpdA
ScaI
BglII
bla
pGBDEL
amdS
NotI
SmaI
XmaI
KpnI

Figure 1

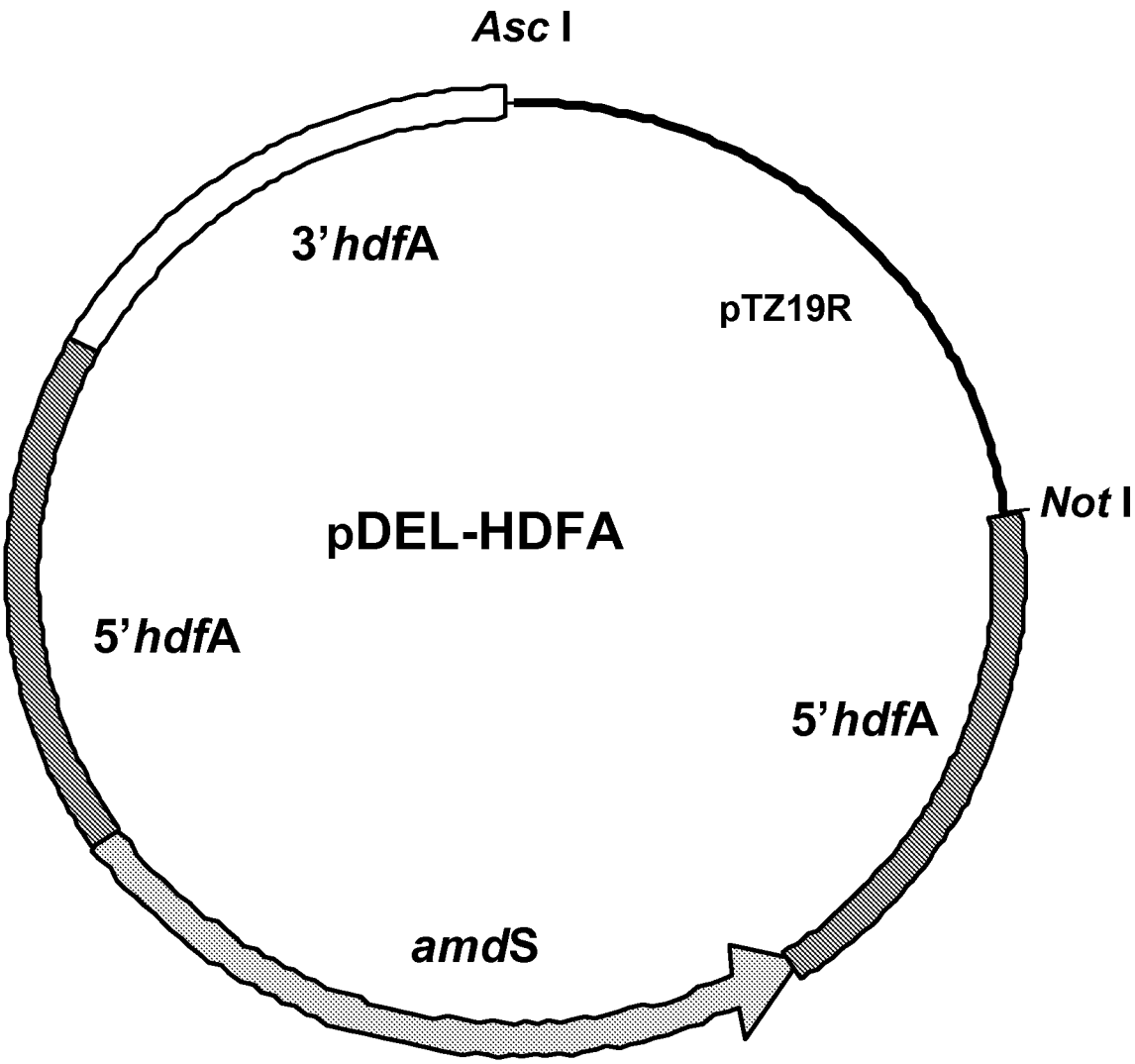


Figure 2

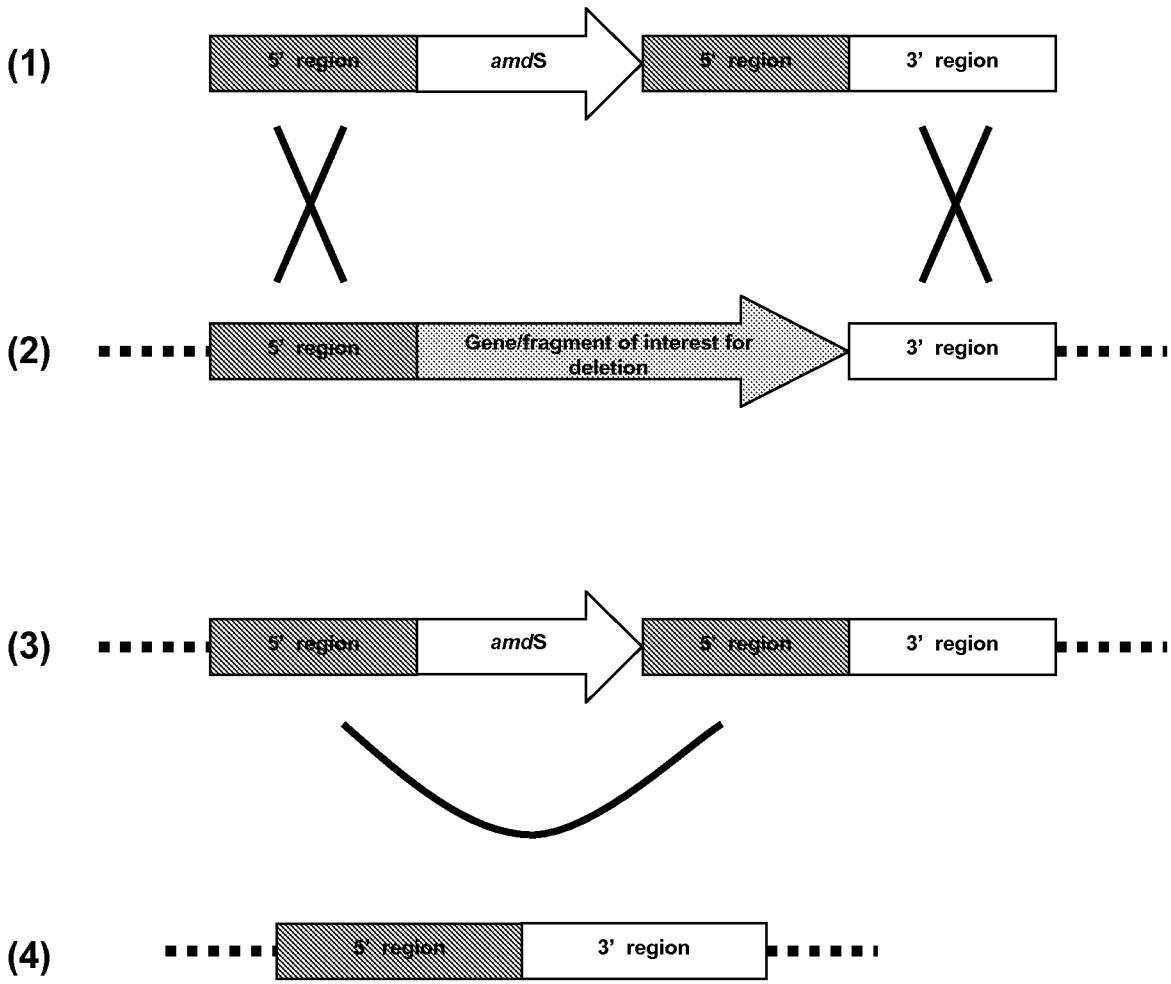

(1)
5' region
amdS
5' region
3' region
(2)
5' region
Gene/fragment of interest for deletion
3' region
(3)
5' region
amdS
5' region
3' region
(4)
5' region
3' region

Figure 3

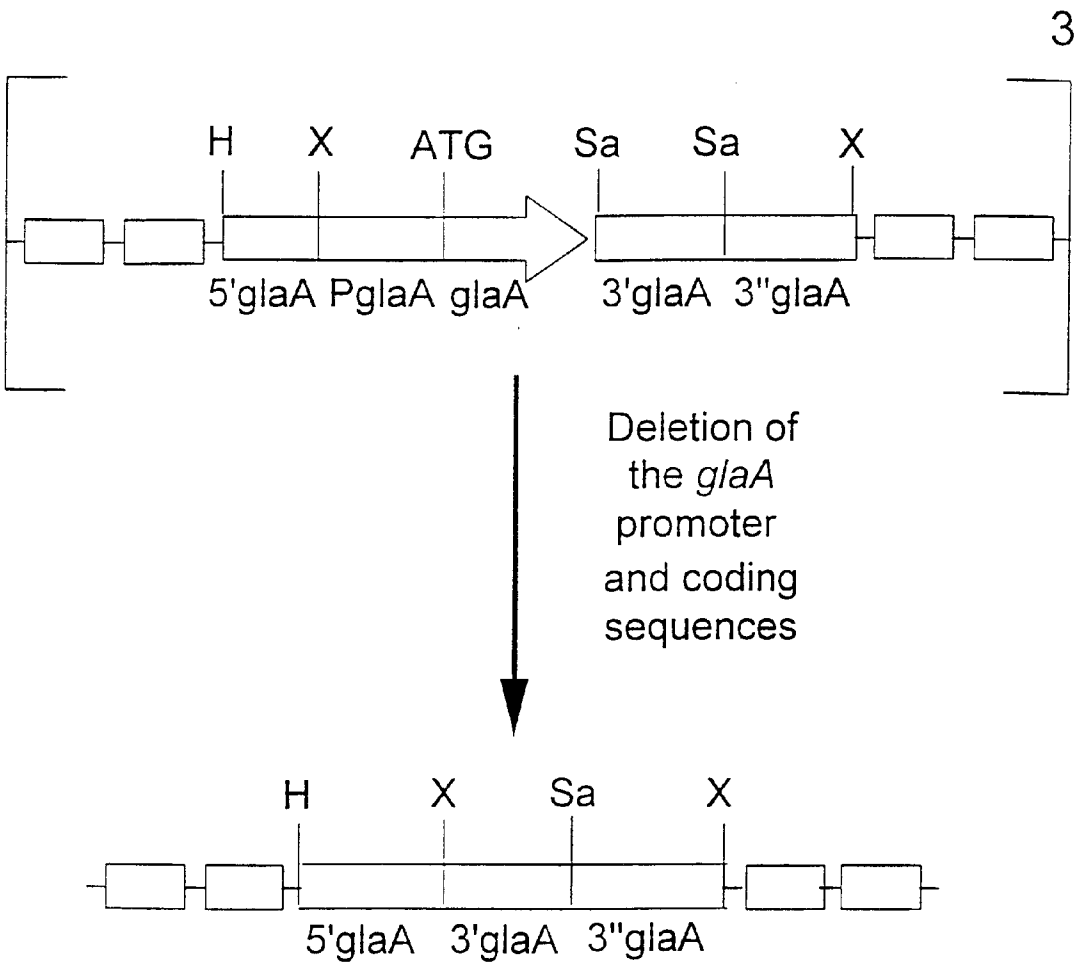

3
H
X
ATG
Sa
Sa
X
5'glaA PglaA glaA
3'glaA 3''glaA
Deletion of the *glaA* promoter and coding sequences
H
X
Sa
X
5'glaA 3'glaA 3''glaA

Figure 4

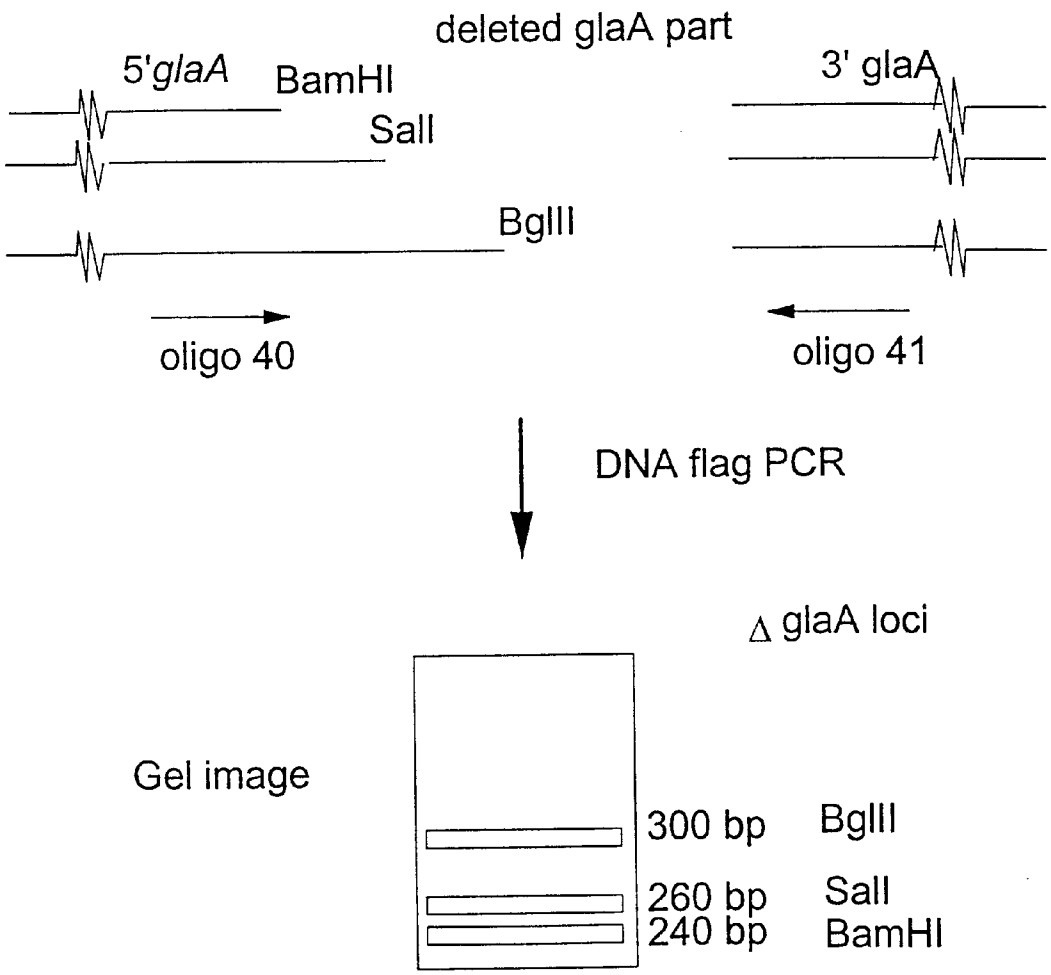

deleted glaA part
5'glaA
BamHI
SalI
BglII
3' glaA
oligo 40
oligo 41
DNA flag PCR
Δ glaA loci
Gel image
300 bp
BglII
260 bp
SalI
240 bp
BamHI

Figure 5

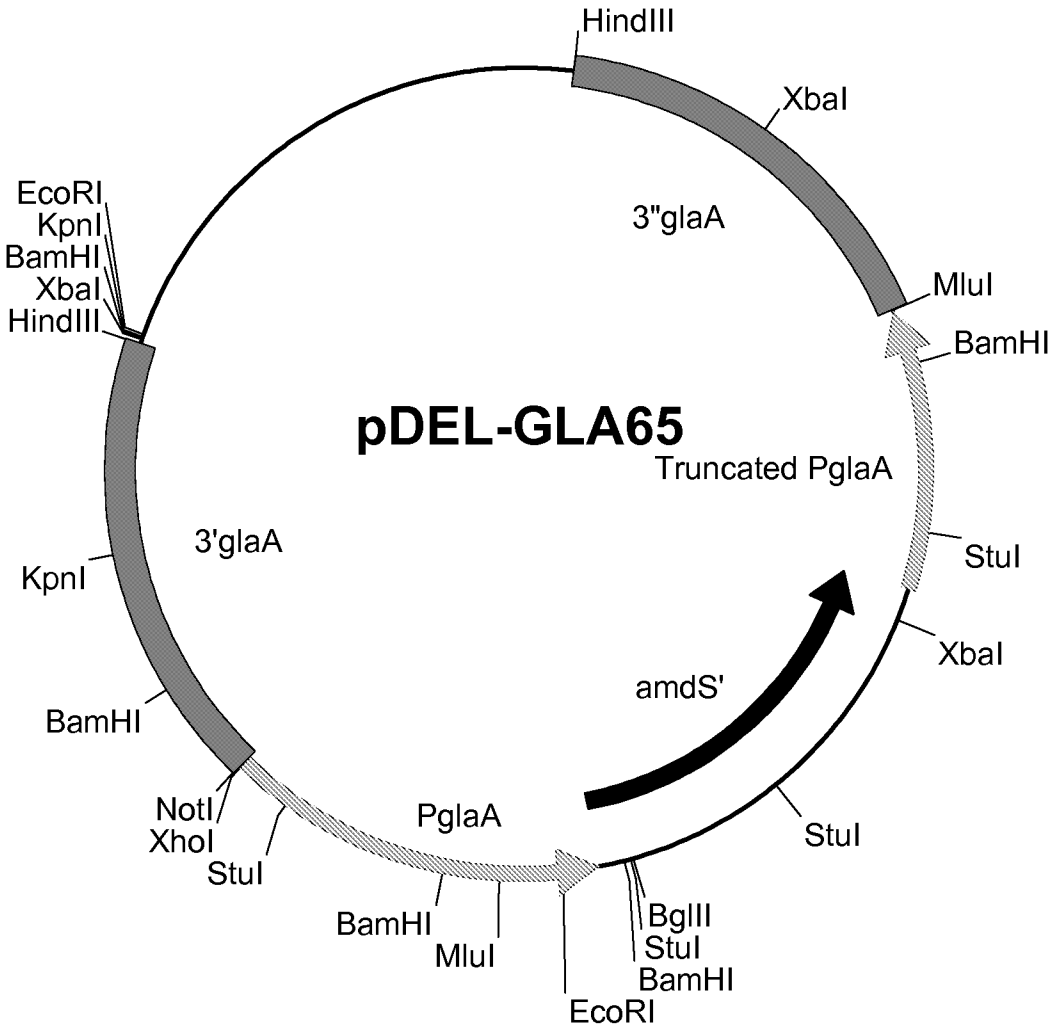

HindIII
XbaI
3"glaA
MluI
BamHI
StuI
XbaI
StuI
BglII
StuI
BamHI
EcoRI
BamHI
MluI
PglaA
NotI
XhoI
StuI
BamHI
KpnI
3'glaA
EcoRI
KpnI
BamHI
XbaI
HindIII
pDEL-GLA65
Truncated PglaA
amdS'

Figure 6

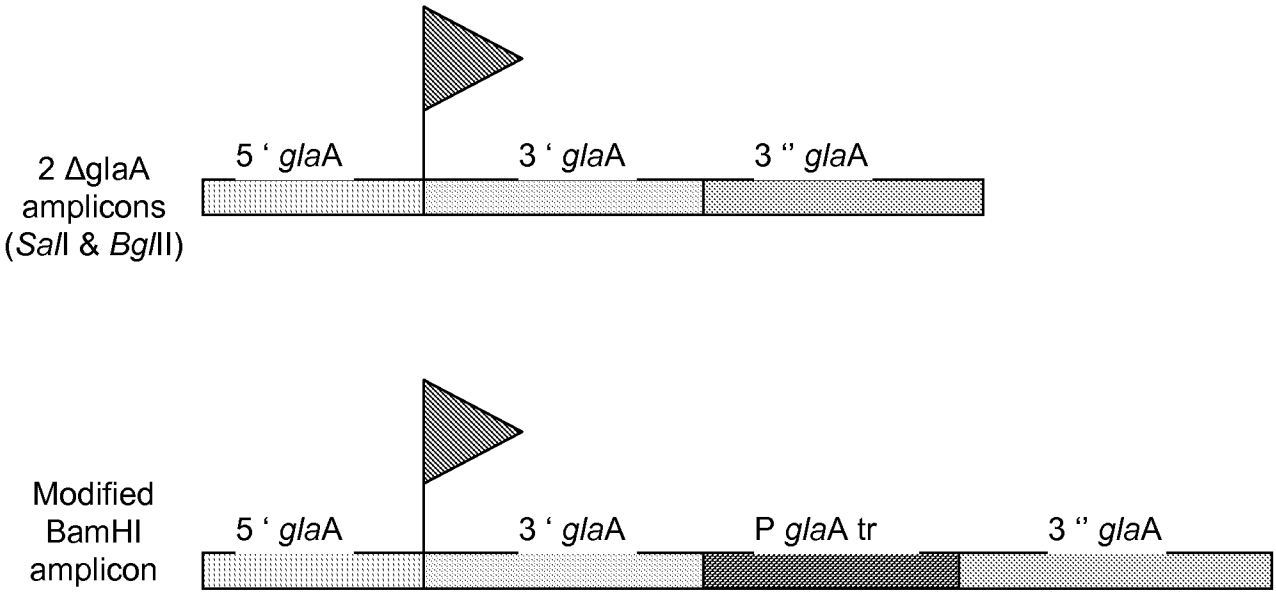

2 ΔglaA amplicons (SalI & BglII)
5 ' glaA
3 ' glaA
3 '' glaA
Modified BamHI amplicon
5 ' glaA
3 ' glaA
P glaA tr
3 '' glaA

Figure 7

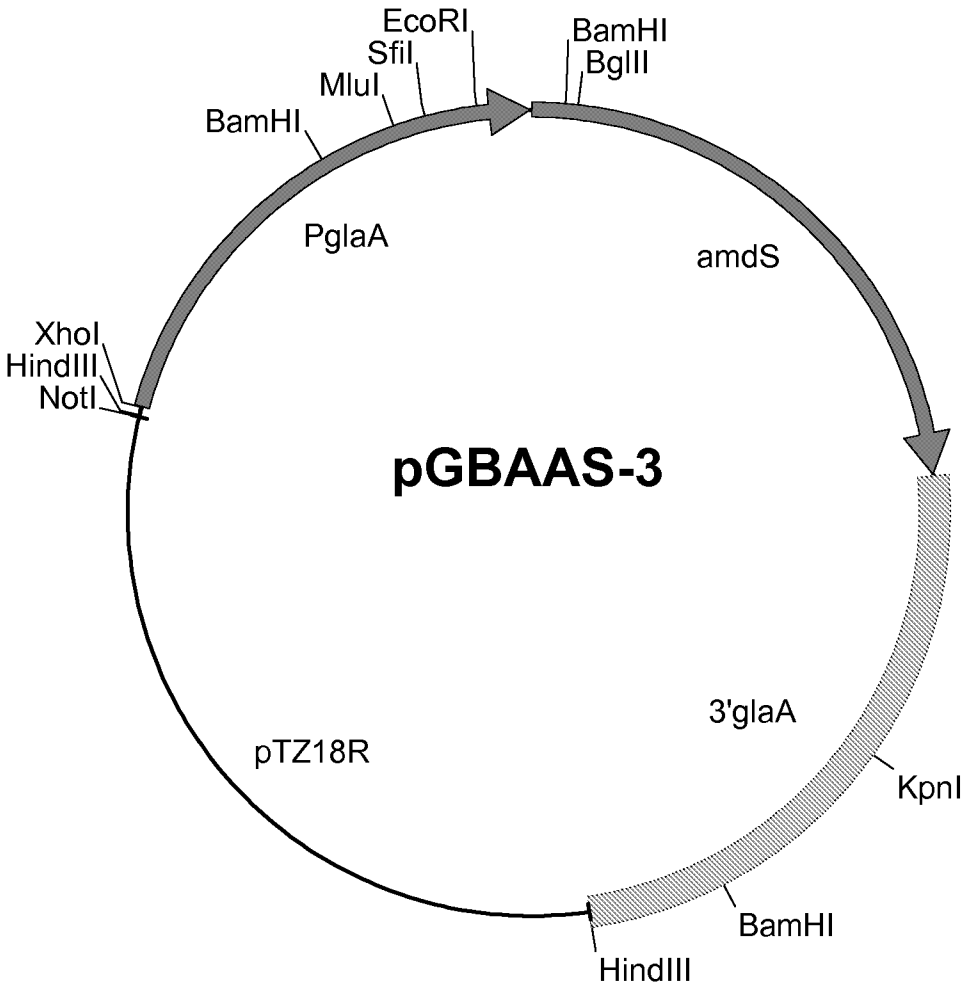

EcoRI
SfiI
MluI
BamHI
BamHI
BglII
PglaA
amdS
XhoI
HindIII
NotI
pGBAAS-3
3'glaA
pTZ18R
KpnI
BamHI
HindIII

Figure 8

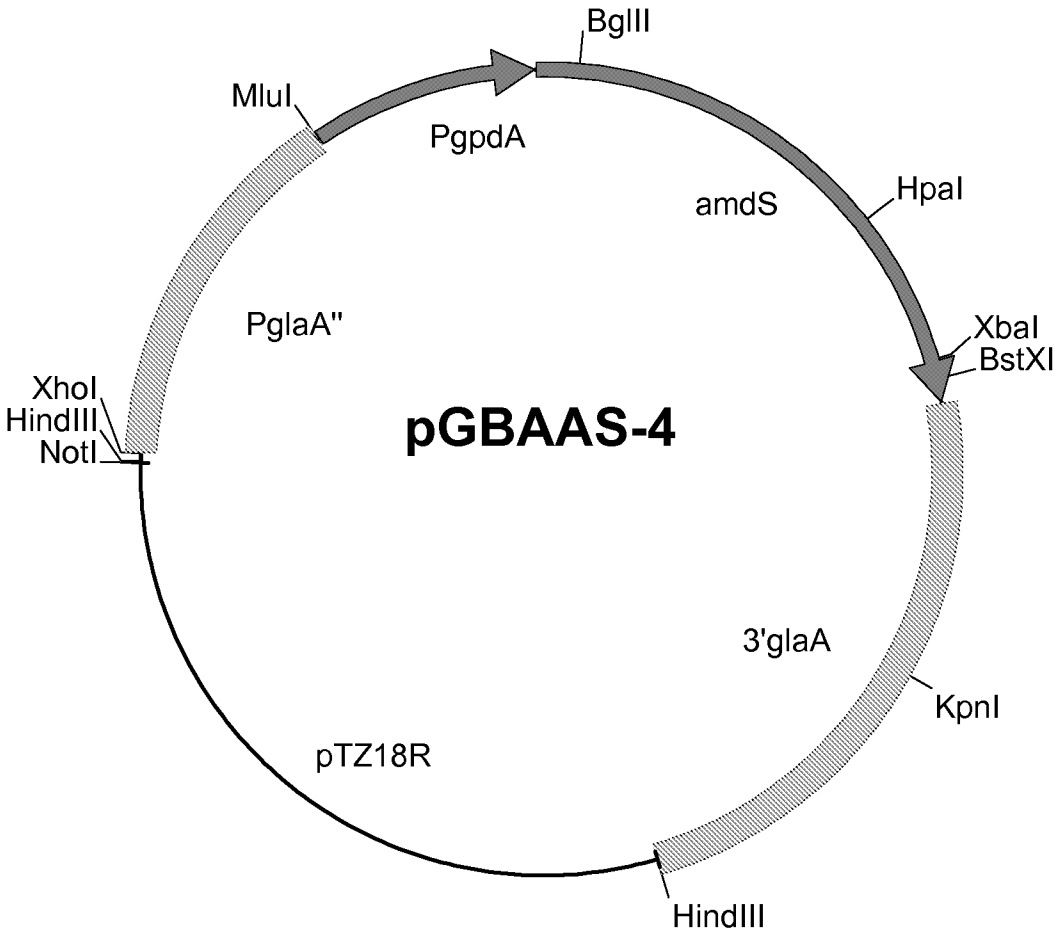

BglII
MluI
PgpdA
amdS
HpaI
PglaA"
XbaI
BstXI
XhoI
HindIII
NotI
pGBAAS-4
3'glaA
KpnI
pTZ18R
HindIII

Figure 9

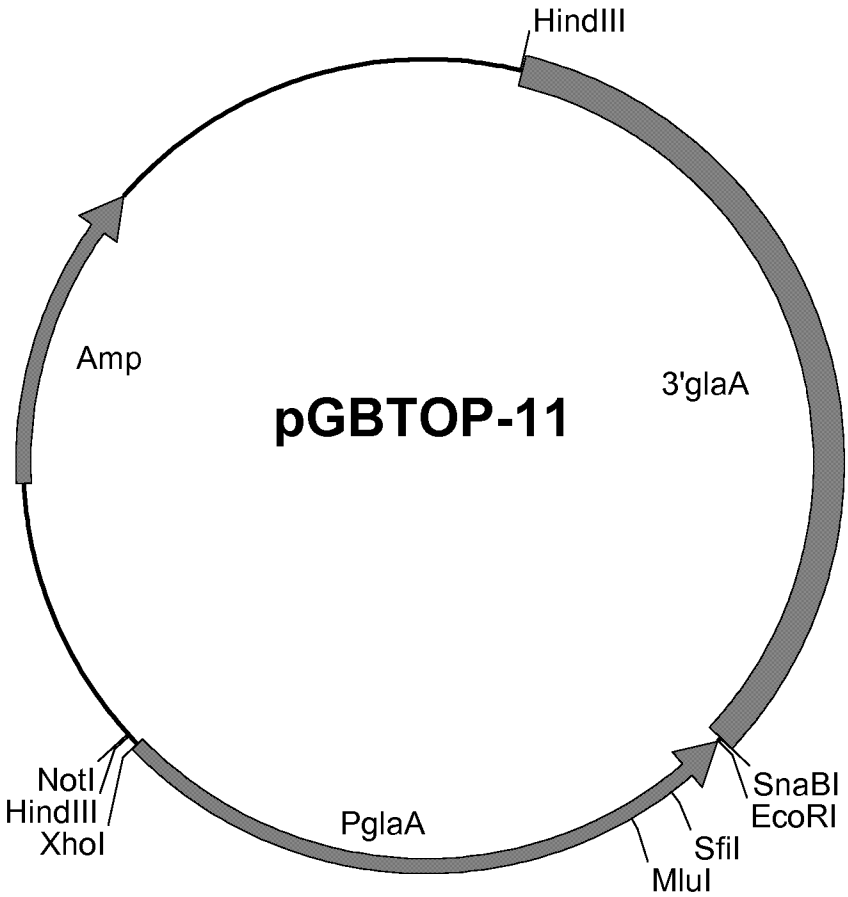

HindIII
Amp
pGBTOP-11
3'glaA
NotI
HindIII
XhoI
PglaA
SfiI
MluI
SnaBI
EcoRI

Figure 10

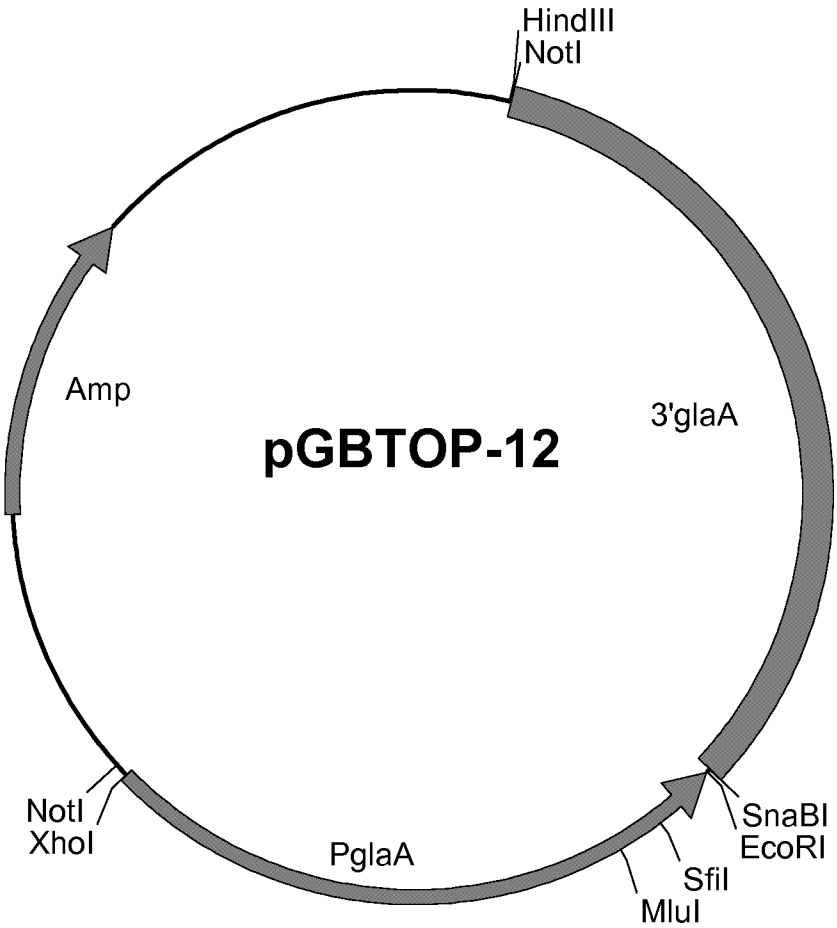

HindIII
NotI
Amp
pGBTOP-12
3'glaA
NotI
XhoI
PglaA
SfiI
MluI
SnaBI
EcoRI

Figure 11

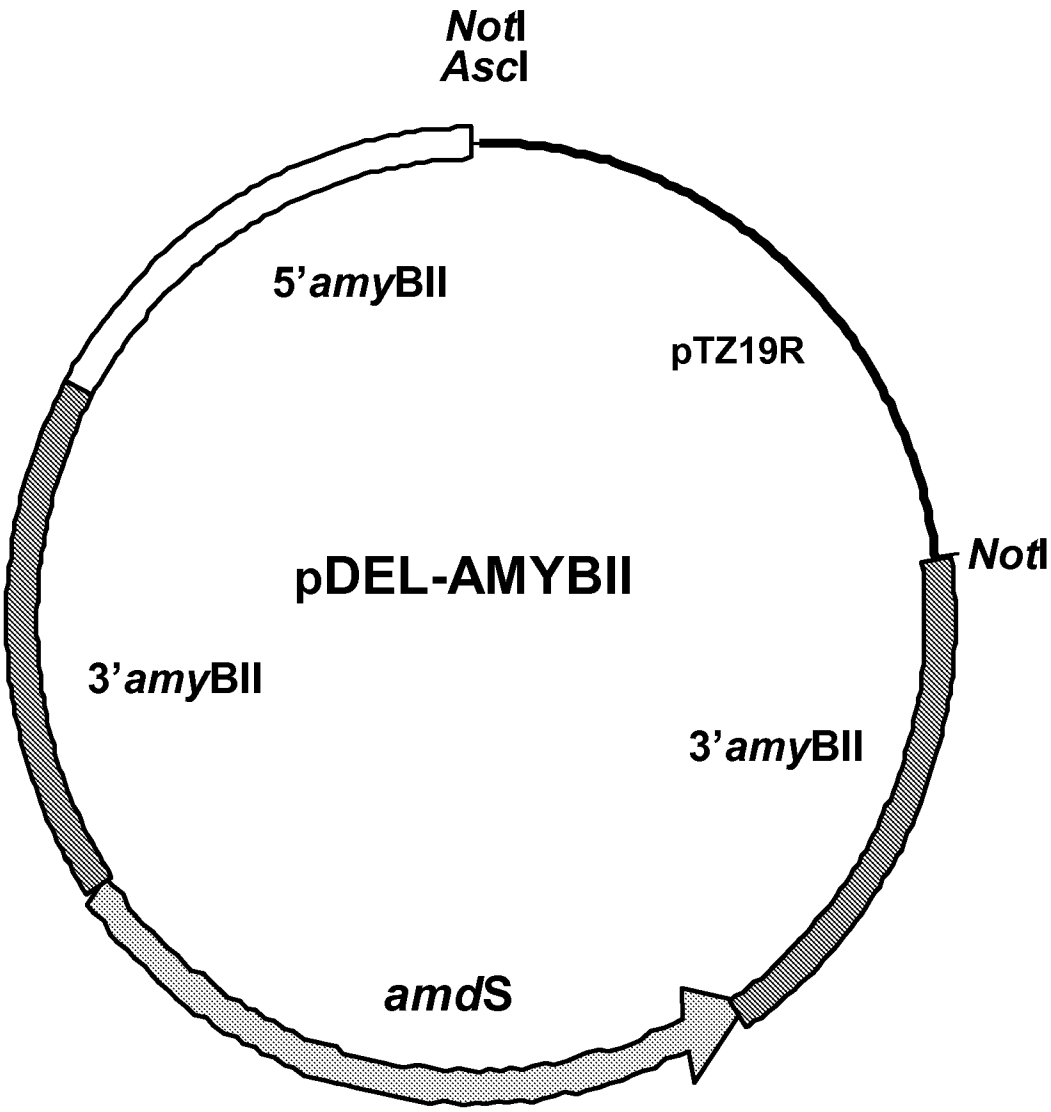

NotI
AscI
5'amyBII
pTZ19R
pDEL-AMYBII
NotI
3'amyBII
3'amyBII
amdS

Figure 12

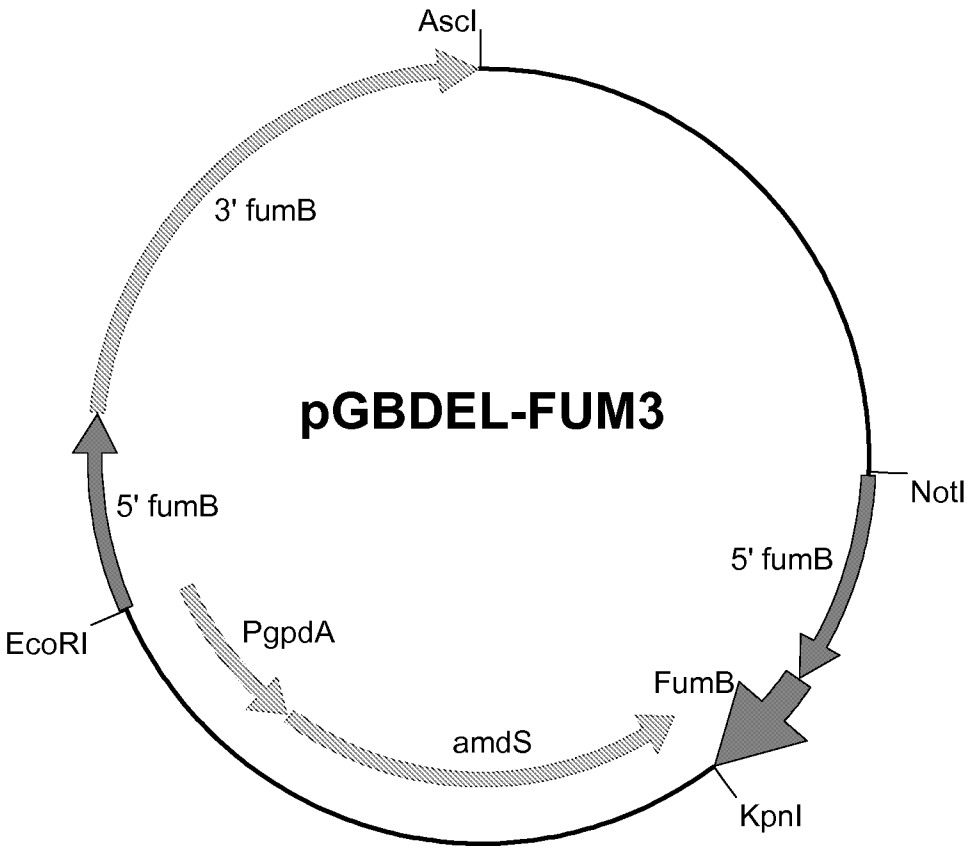

AscI
3' fumB
pGBDEL-FUM3
5' fumB
NotI
5' fumB
EcoRI
PgpdA
FumB
amdS
KpnI

Figure 13

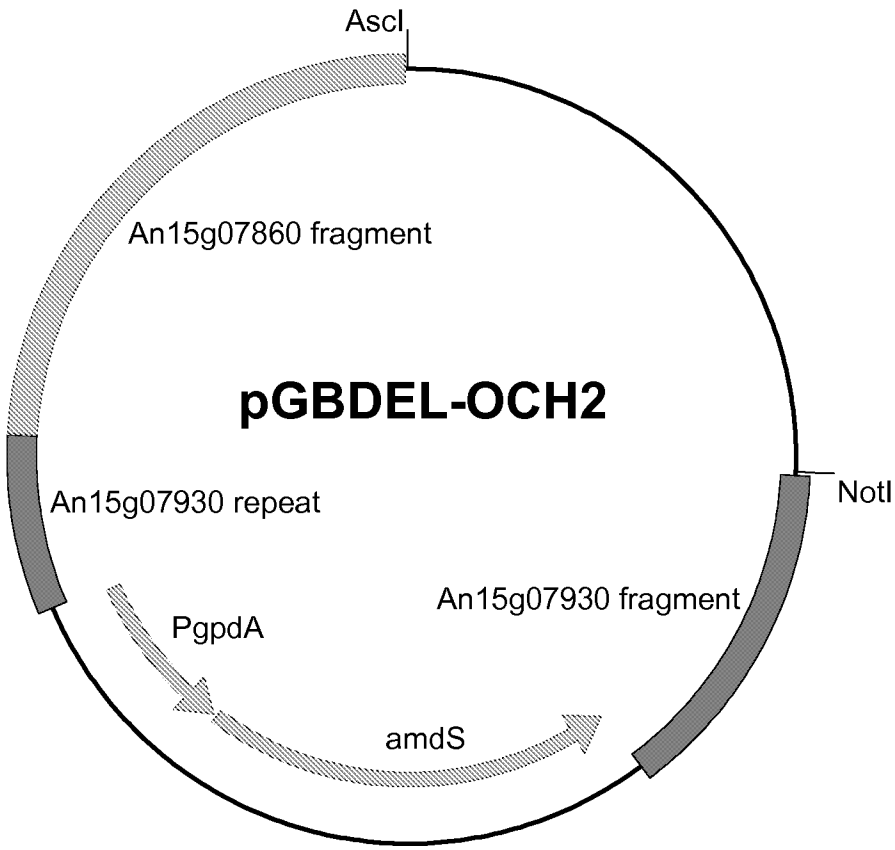

AscI
An15g07860 fragment
pGBDEL-OCH2
An15g07930 repeat
NotI
An15g07930 fragment
PgpdA
amdS

Figure 14

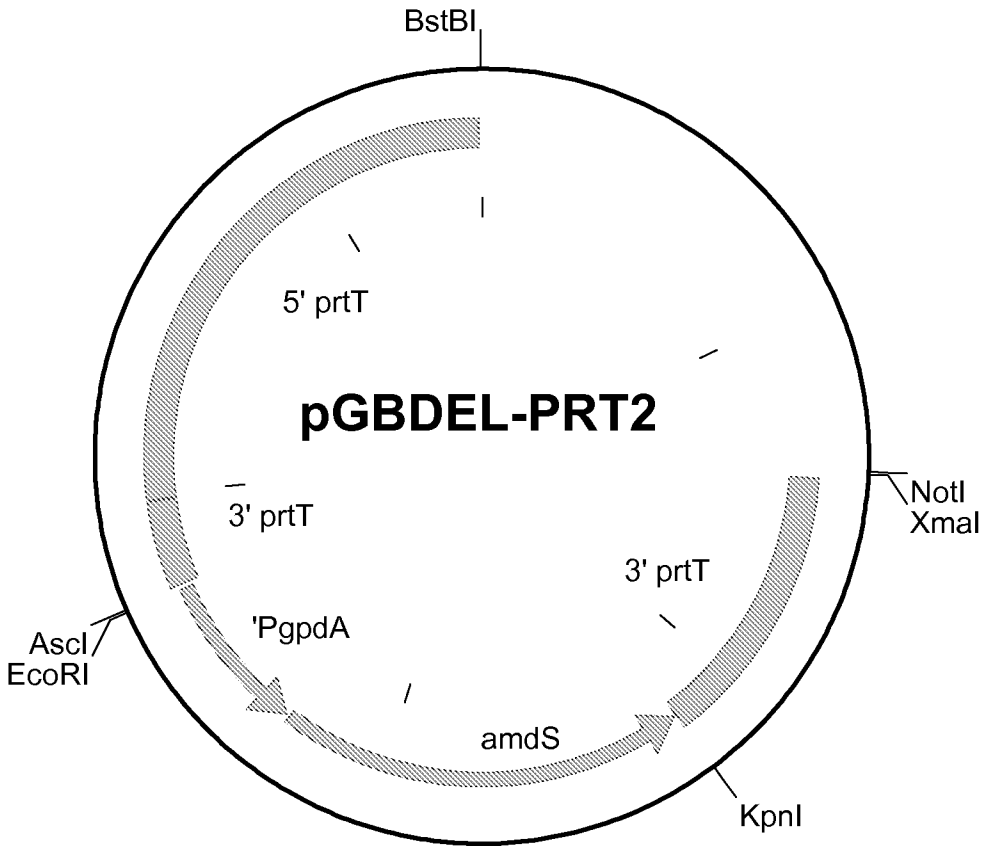

BstBI
5' prtT
pGBDEL-PRT2
NotI
XmaI
3' prtT
3' prtT
'PgpdA
AscI
EcoRI
amdS
KpnI

Figure 15

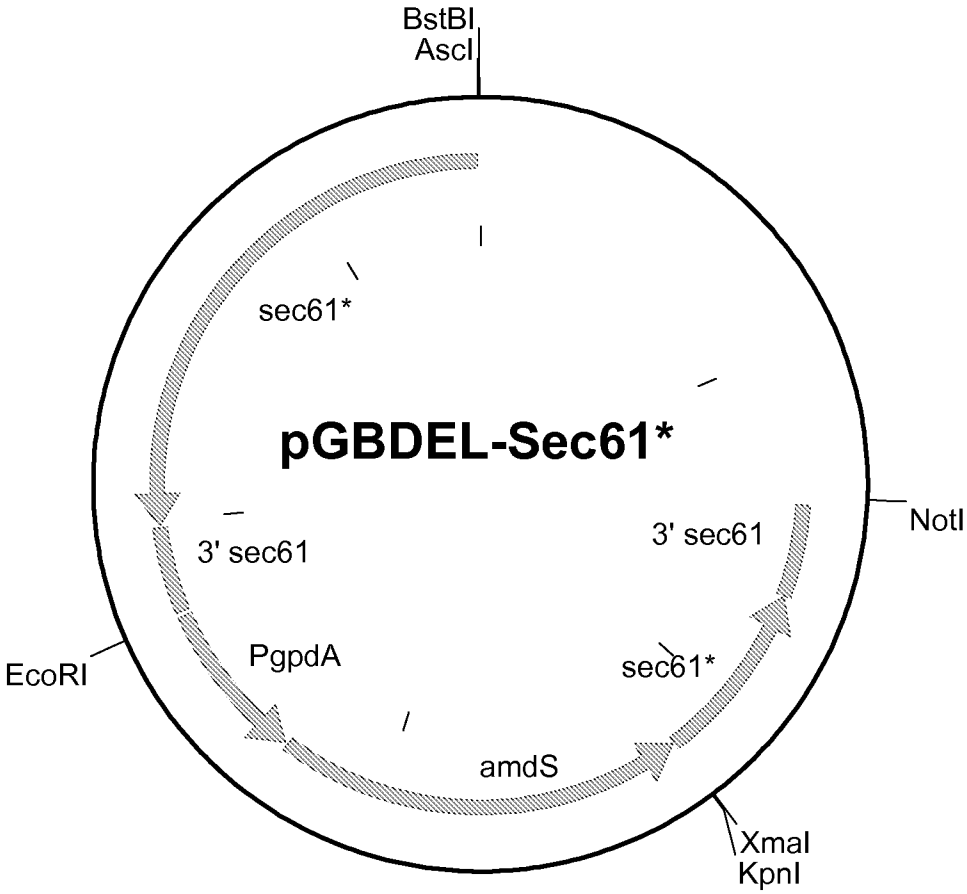

BstBI
AscI
sec61*
pGBDEL-Sec61*
NotI
3' sec61
3' sec61
EcoRI
PgpdA
sec61*
amdS
XmaI
KpnI

Figure 16

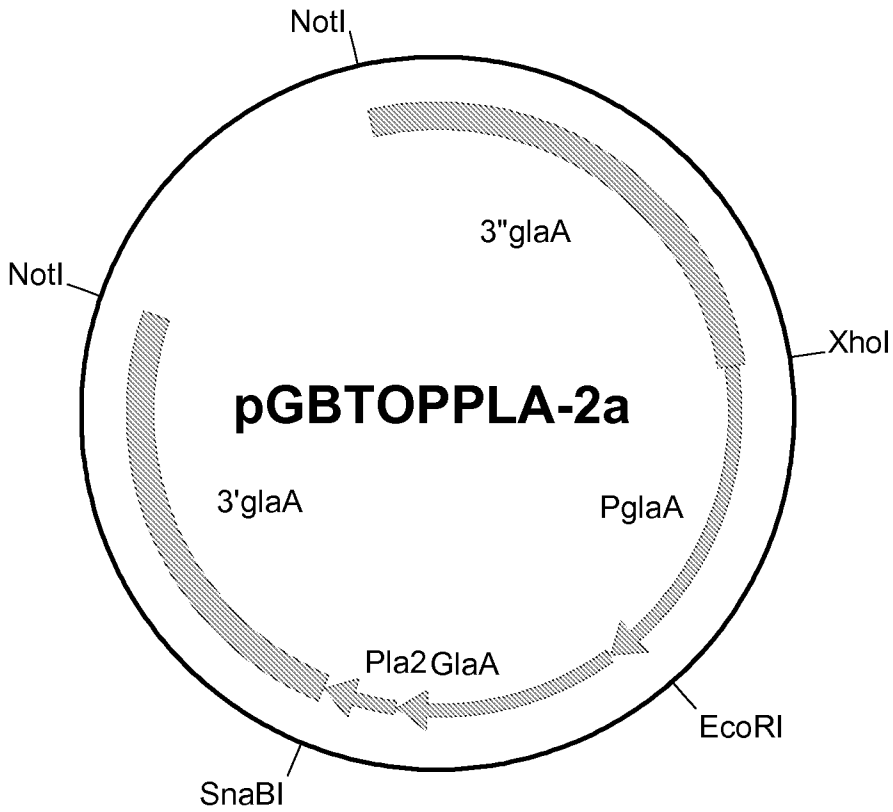

NotI
3"glaA
NotI
XhoI
pGBTOPPLA-2a
3'glaA
PglaA
Pla2GlaA
EcoRI
SnaBI

Figure 17

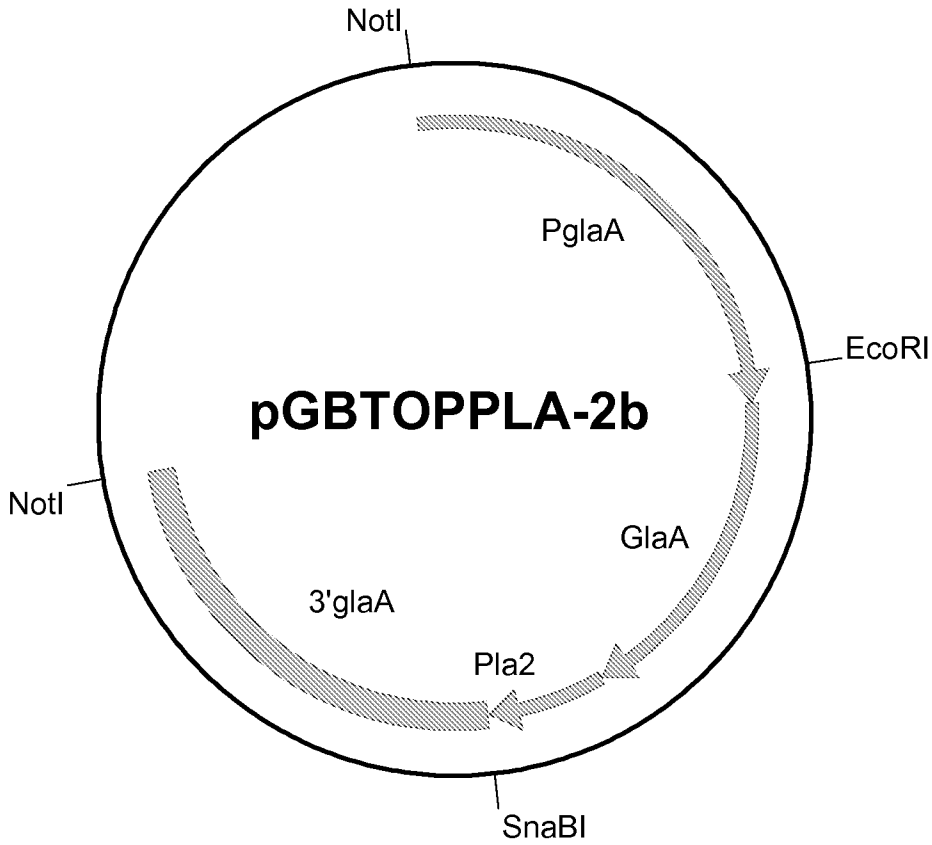


Figure 18

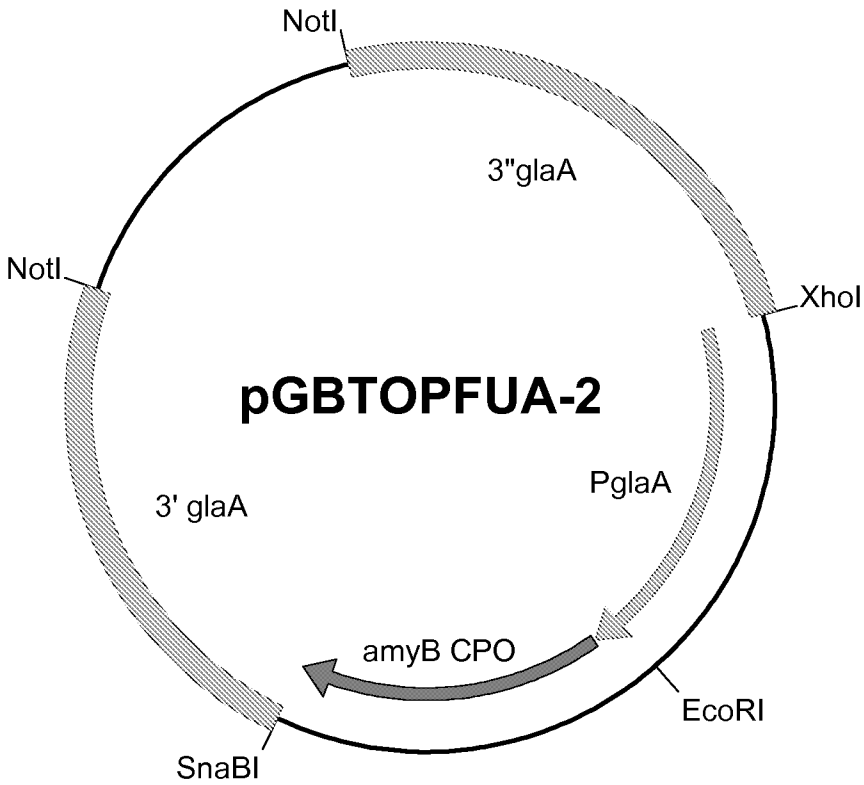

NotI
3"glaA
NotI
XhoI
pGBTOPFUA-2
PglaA
3' glaA
amyB CPO
EcoRI
SnaBI

Figure 19

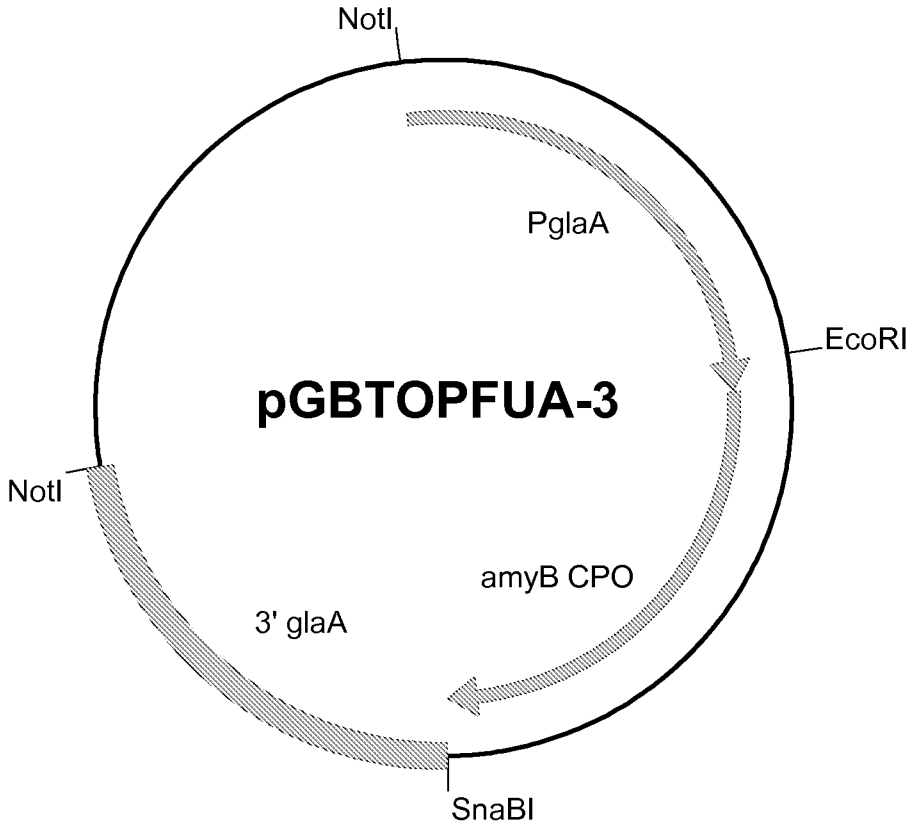


Figure 20

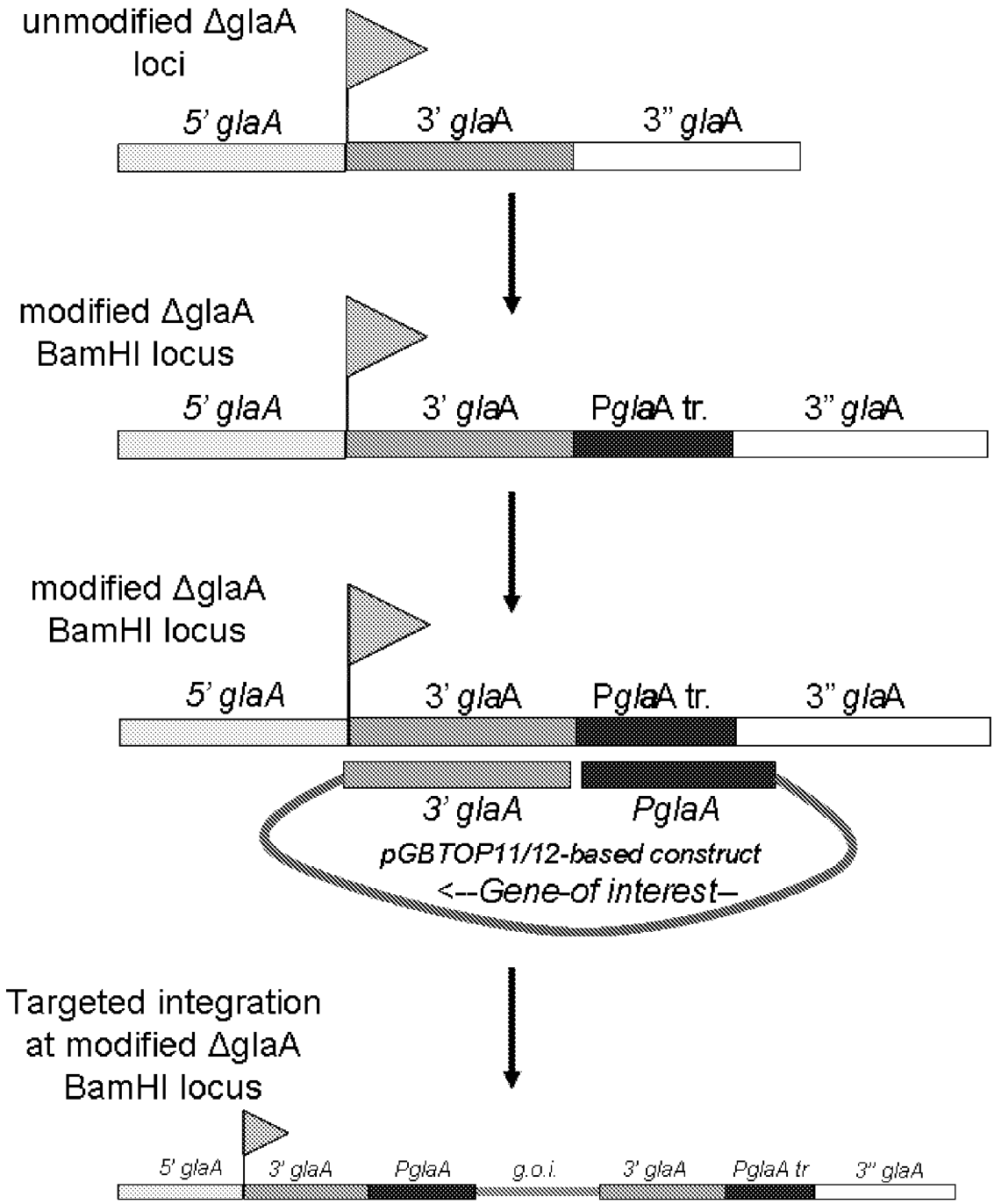

unmodified ΔglaA loci
5' glaA
3' glaA
3" glaA
modified ΔglaA BamHI locus
5' glaA
3' glaA
PglaA tr.
3" glaA
modified ΔglaA BamHI locus
5' glaA
3' glaA
PglaA tr.
3" glaA
3' glaA
PglaA
pGBTOP11/12-based construct
<--Gene-of interest--
Targeted integration at modified ΔglaA BamHI locus
5' glaA
3' glaA
PglaA
g.o.i.
3' glaA
PglaA tr
3" glaA

Figure 21

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2010/059390

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/81
ADD. C12N15/00 C12N1/16 C12N15/56 C12N9/60

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

B. FIELDS SEARCHED

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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2008/098933 A1 (DSM IP ASSETS BV [NL]; SAGT CORNELIS MARIA JACOBUS [NL]; VAN PEIJ NOEL) 21 August 2008 (2008-08-21) the whole document	1-22, 25-28
X	WO 2007/115886 A1 (DSM IP ASSETS BV [NL]; BERG VAN DEN MARCO ALEXANDER [NL]; KERKMAN RICH) 18 October 2007 (2007-10-18) the whole document	1-8, 11-24
X	WO 98/46774 A1 (GIST BROCADES BV [NL]; SWINKELS BART WILLEM [NL]; OOIJEN ALBERT JOHANN) 22 October 1998 (1998-10-22) the whole document	1-5, 14-24

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 August 2010	02/09/2010

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2010/059390

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. (means)

   [ ] on paper

   [X] in electronic form

   b. (time)

   [X] in the international application as filed

   [ ] together with the international application in electronic form

   [ ] subsequently to this Authority for the purpose of search

2. [ ] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 1-24(partially)

Present claims 1 and 14 relate to a product and a method defined (inter alia) by reference to the following unusual parameter: [...] where the adapted substantially homologous DNA domain originates from has a gene conversion frequency that is (at least 10%) higher than one of the other of the at least two substantially homologous DNA domains.[...] The use of this unusual parameter in the present context is considered to lead to a lack of clarity, because the claim does not clearly identify the products encompassed by it as the parameter cannot be clearly and reliably determined by indications in the description or by objective procedures which are usual in the art. This makes it impossible to compare the claim with the prior art. As a result, the application does not comply with the requirement of clarity under Article 84 EPC. Moreover, the application provides example for only one organisms where this unusual parameter could probably be validated, whereas the claim encompasses all host cells, which produced a further problem under Art.83 and 84 EPC. The lack of clarity, support and disclosure is such that a meaningful search of the whole claimed subject-matter of said claim could not be carried out (Rule 63 EPC). The extent of the search was consequently limited to the examples clearly defined in, and supported and disclosed by, the description.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.2), should the problems which led to the Article 17(2) declaration be overcome.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2010/059390

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

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

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

2. ☒ Claims Nos.: 1-24(partially)
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

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

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

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

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

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

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

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

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2010/059390

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008098933 A1	21-08-2008	AU 2008214663 A1	21-08-2008
		CA 2677568 A1	21-08-2008
		CN 101784666 A	21-07-2010
		EA 200901110 A1	30-04-2010
		EP 2115145 A1	11-11-2009
		JP 2010517587 T	27-05-2010
		US 2010112638 A1	06-05-2010
WO 2007115886 A1	18-10-2007	AT 460489 T	15-03-2010
		CN 101421401 A	29-04-2009
		DK 2004827 T3	14-06-2010
		EP 2004826 A1	24-12-2008
		EP 2004827 A1	24-12-2008
		WO 2007115887 A1	18-10-2007
		US 2009124014 A1	14-05-2009
		US 2009170206 A1	02-07-2009
WO 9846774 A1	22-10-1998	AT 398181 T	15-07-2008
		AU 8013798 A	11-11-1998
		DK 0973918 T3	22-09-2008
		EP 0973918 A1	26-01-2000
		ES 2308805 T3	01-12-2008
		JP 2002514077 T	14-05-2002