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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 August 2015 (20.08.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/120542 A1

(51) International Patent Classification:
C07K 19/00 (2006.01) *C12N 15/63* (2006.01)
C07K 14/025 (2006.01) *C12N 15/85* (2006.01)
C07K 14/16 (2006.01) *C12N 15/87* (2006.01)
C07K 14/435 (2006.01) *C12N 5/10* (2006.01)
C12N 1/21 (2006.01) *C12P 21/02* (2006.01)
C12N 15/62 (2006.01)

(21) International Application Number:
PCT/CA2015/000084

(22) International Filing Date:
12 February 2015 (12.02.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/940,274	14 February 2014 (14.02.2014)	US
61/940,258	14 February 2014 (14.02.2014)	US
62/013,852	18 June 2014 (18.06.2014)	US

(71) Applicant: **SYMVIVO CORPORATION** [CA/CA]; 205-4475 Wayburne Drive, Burnaby, British Columbia V5G 4X4 (CA).

(72) Inventors: **GRAVES, Herbert Alexander**; 4890 Quebec Street, Vancouver, British Columbia V5V 3M3 (CA). **FOX, Mark Andrew**; 2005-833 Homer Street, Vancouver, British Columbia V6V 0H4 (CA).

(74) Agent: **SECHLEY, Konrad**; Gowling Lafleur Henderson LLP, 550 Burrard Street, Suite 2300, Bentall 5, Vancouver, British Columbia V6C 2B5 (CA).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

[Continued on next page]

(54) Title: HYBRID PROTEINS AND USES THEREOF

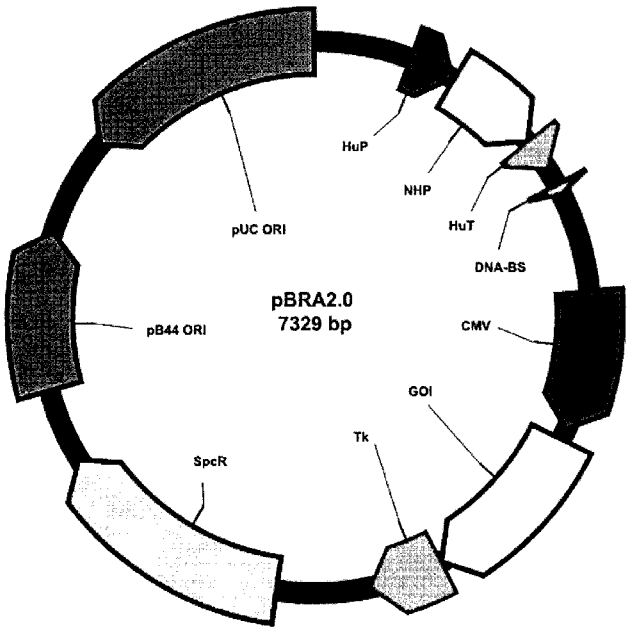


FIG. 1

pBRA 2.0-SHT

(57) Abstract: There are disclosed hybrid proteins comprising at least one signal sequence; at least one DNA binding domain; and at least one cell penetrating peptide (CPP) domain. In embodiments the CPP domain is a TAT domain, and the DNA binding domain is a HU domain. There is also disclosed the use of the hybrid proteins to introduce exogenous DNA into target cells, and methods for introducing exogenous DNA into target cells using the hybrid proteins.

WO 2015/120542 A1

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

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

HYBRID PROTEINS AND USES THEREOF

PRIORITY CLAIMS

This application claims priority under 35USC§119(e) of US provisional patent application number 61/940,274, filed February 14, 2014; US provisional patent application number 61/940,258 filed February 14, 2014; and US provisional patent application number 62/013,852 filed June 18, 2014. The specifications of which are hereby incorporated by reference wherever permissible by.

BACKGROUND

1. Field

The Subject matter disclosed generally relates to hybrid proteins and the use of the hybrid proteins to transform prokaryotic and eukaryotic cells.

2. RELATED ART

A variety of vectors and methods are known in the art for propagating nucleic acid sequences in bacteria and for introducing nucleic acid sequences into different types of bacteria and into eukaryotic cells. The following publications are of note:

Salomone, F., et al., "A novel cell-penetrating peptide with membrane disruptive properties for efficient endosomal escape" (2010), Journal of controlled release 163 (293-303).

Stentz, R. et al., "Controlled release of protein from viable Lactococcus cells" (2010) Applied and Environmental Microbiology 76, 3026-3031.

Christy, B., and Nathans, D. "DNA binding site of the growth factor-inducible protein Zif268" (1989) 86 Proc. Nat. Acad. Sci. 8737-8741.

Khokhlova, E.V., et al., "Bifidobacterium Longum Modifed recombinant HU protein as vector for nonviral delivery of DNA to HEK293 Human cell culture" (2011) 151 Bulletin of Experimental Biology and Medicine 717-721.

SUMMARY

In an embodiment there is disclosed a hybrid protein comprising: at least one signal sequence; at least one DNA binding domain; and at least one cell penetrating peptide (CPP) domain.

In alternative embodiments, at least one CPP domain is a TAT domain, a VP22 domain, an Antp domain, a Rev domain, a P-beta (gp41-SV40) domain, a Transportan (Galanin-mastoparan) domain or a Pep-1 (Trp-rich motif SV40) domain.

In alternative embodiments, at least one secretion signal sequence is selected from the group consisting of an alpha amylase signal sequence, a truncated alpha amylase signal sequence and an alpha arabinosidase signal sequences.

In alternative embodiments, at least one DNA binding domain is a sequence specific DNA binding domain.

In alternative embodiments, at least one DNA binding domain comprises at least one domain selected from the group consisting of: a Zinc finger DNA binding domain, a homeobox DNA binding domain, a MerR DNA binding domain, and a HU DNA binding domain.

In alternative embodiments, the hybrid protein may comprise an alpha arabinosidase signal sequence, a TAT domain and a HU DNA binding domain.

In an embodiment there is disclosed a method for transforming a target cell with a desired DNA, the method comprising the step of: contacting the target cell with a protein-DNA complex comprising the desired DNA and the hybrid protein disclosed herein.

In alternative embodiments, the DNA is a plasmid.

In alternative embodiments, the plasmid comprises an expression cassette for expressing the hybrid protein in a gram positive bacterium.

In alternative embodiments, at least one DNA binding domain is a sequence specific DNA binding domain and wherein the plasmid comprises at least one nucleotide sequence for binding to at least one sequence specific DNA binding domain.

In alternative embodiments, at least one CPP sequence has at least 90% amino acid sequence identity over a sequence of at least 10 contiguous amino acids to SEQ ID NO: 11, 12, 13, 14, 15, 16, or 17.

In alternative embodiments, at least one signal sequence has at least 90% amino acid sequence identity over a sequence of at least 10 contiguous amino acids to SEQ ID NO: 19, 21 or 23.

In alternative embodiments, at least one DNA binding domain has at least 90% amino acid sequence identity over at least 10 amino acids to SEQ. ID. NO. 44, 45 or 46.

In alternative embodiments, the cell is a gram positive bacterium.

In alternative embodiments, the cell is a mammalian cell.

In alternative embodiments, the method may further comprise the step of contacting the DNA with the hybrid protein to form the protein-DNA complex.

In an embodiment there is disclosed a cell transformed with an exogenous DNA using the hybrid protein disclosed herein, wherein the cell is a gram positive bacterium or is a mammalian cell.

In an embodiment there is disclosed the use of the hybrid protein disclosed herein to transform a gram positive bacterium or a mammalian cell.

In an embodiment there is disclosed the use of the hybrid protein disclosed herein to transform a cell selected from the group consisting of: *Staphylococcus, streptococcus, Bifidobacterium, lactococcus, lactobacillus, clostridium.*

In an embodiment there is disclosed a kit for transforming a target cell with a desired DNA, the kit comprising a quantity of the hybrid protein disclosed herein, and instructions to: form a complex of the hybrid protein with the desired DNA; and contact the complex with the target cell.

In an embodiment there is disclosed a gram positive bacterium able to synthesize the novel hybrid protein disclosed herein.

In an embodiment there is disclosed a hybrid protein comprising: at least one CPP domain selected from the group consisting of a TAT domain, a VP22 domain, an Antp domain, a Rev domain, a P-beta (gp41-SV40) domain, a Transportan (Galanin-mastoparan) domain or a Pep-1 (Trp-rich motif SV40) domain; at least one DNA binding domain selected from the group consisting of a Zinc finger DNA binding domain, a homeobox DNA binding domain, a MerR DNA

binding domain, and a HU DNA binding domain; and at least one signal sequence selected from the group consisting of alpha amylase signal sequence, a truncated alpha amylase signal sequence and an alpha arabinosidase signal sequence.

Features and advantages of the subject matter hereof will become more apparent in light of the following detailed description of selected embodiments, as illustrated in the accompanying figures. As will be realized, the subject matter disclosed and claimed is capable of modifications in various respects, all without departing from the scope of the claims. Accordingly, the drawings and the description are to be regarded as illustrative in nature, and not as restrictive and the full scope of the subject matter is set forth in the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG.1. shows the structure of a vector according to a first embodiment of the present subject matter and identified as pBRA2.0 SHT.

FIG.2. shows the structure of a vector according to a second embodiment of the present subject matter and identified as pFRG1.5-SHT.

FIG. 3 shows a gel shift assay demonstrating the binding of SHT and SZT hybrid proteins with pBRA2.0 SHT. SHT and SZT interact with pBRA2.0 and result in gel-shift with increasing concentrations of peptide.

FIG. 4 shows the results of transforming *E. coli* with pBRA2.0-SHT vector according to an embodiment of the present subject matter, to confirm the effectiveness of the *E. coli* (pUC) origin of replication.

FIG. 5 shows the results of transfecting HEK-293 and HeLa cells with pBRA2.0 SHT vector according to an embodiment of the present subject matter.

FIG. 6 shows the secretion of pBRA2.0 SHT from *Bifidobacterium longum* cells hosting the vector.

FIG. 7 shows digestion products of pBRA2.0 SHT from cell supernatant of pBRA2.0 SHT infected *Bifidobacterium longum* cells.

FIG. 8 shows the visualisation and digestion of concentrate cell supernatants showing secretion of pBRA2.0 SHT and its characterisation.

FIG. 9 shows immunoblotting analysis of the SZT hybrid protein demonstrating the expression of the SZT protein in *Bifidobacterium*.

FIG. 10 shows the transfection of mammalian cells by protein complexes comprising pBRA2.0-SHT or pBRA2.0-SZT, and the expression of the cargo GFP sequences in the mammalian cell lines. Therapeutic molecules pBRA2.0-SHT and pBRA2.0-SZT complexes can transfect and express GFP in mammalian cell lines.

DETAILED DESCRIPTION OF EMBODIMENTS

The following sequence listings are presented herein and form a part of this disclosure:

SEQ ID NO: 1 is the sequence of pBRA2.0 SHT wherein the eukaryotic expression cassette encodes GFP protein and the prokaryotic expression cassette encodes the SHT protein. Source: Artificial.

SEQ ID NO: 2. is the sequence of pFRG1.5-SHT wherein the eukaryotic expression cassette comprises the green fluorescent protein (GFP) gene. Source: Artificial.

SEQ ID NO: 3 is the nucleotide sequence encoding the Lac I DNA binding domain according to embodiments. Source: E. coli.

SEQ ID NO: 4 is the nucleotide sequence encoding the HU DNA binding domain according to embodiments. Source: Bifidobacterium.

SEQ ID NO: 5 is the nucleotide sequence encoding the Mer R DNA binding domain according to embodiments. Source: Bifidobacterium.

SEQ ID NO: 6 is the nucleotide sequence encoding the Zinc finger DNA binding domain according to embodiments. Source: Artificial.

SEQ ID NO: 7 is the nucleotide sequence encoding the SMT hybrid protein Source: Artificial.

SEQ ID NO: 8 is the nucleotide sequence encoding the SHT hybrid protein. Source: Artificial.

SEQ. ID. N0. 9 is the nucleotide sequence encoding the SLT hybrid protein. Source: Artificial.

SEQ ID NO: 10 is the nucleotide sequence encoding the SZT hybrid protein. Source: Artificial.

SEQ ID NO: 11 is the amino acid sequence of the Trans-Activator of Transcription (Tat) transduction domain [HIV].

SEQ ID NO: 12 is the amino acid sequence of the Antennapedia (Antp) transduction domain Source: Drosophila Melanogaster.

SEQ ID NO: 13 is the amino acid sequence of the HIV Rev transduction domain. Source: Human HIV virus.

SEQ ID NO: 14 is the amino acid sequence of the herpes simplex virus VP22 transduction domain. Source: Human HSV virus.

SEQ ID NO: 15 is the amino acid sequence of the P-beta MPG (gp41-SV40) transduction domain. Source: SV40.

SEQ ID NO: 16 is the amino acid sequence of the Transportan (Galanin-mastoparan) transduction domain Source: Eukaryote, species unknown.

SEQ ID NO: 17 is the amino acid sequence of the Pep-1 (Trp-rich motif-SV40) transduction domain. Source: SV40

SEQ ID NO: 18 is the DNA sequence encoding the alpha amylase signal sequence comprising a cleavage site. Source: Bifidobacterium.

SEQ ID NO: 19 is the 46 amino acid sequence encoding the alpha amylase signal sequence including a putative cleavage site. Source: Bifidobacterium.

SEQ ID NO: 20 is the DNA sequence encoding the cleaved alpha amylase signal sequence. Source: Bifidobacterium.

SEQ ID NO: 21 is the 44 amino acid sequence of the cleaved alpha amylase signal sequence. Source: Bifidobacterium.

SEQ ID NO: 22 is the DNA sequence encoding the alpha arabinosidase signal sequence. Source: Bifidobacterium.

SEQ ID NO: 23 is the amino acid sequence for the alpha arabinosidase signal sequence. Source: Bifidobacterium.

SEQ ID NO: 24 is the amino acid sequence of the hybrid protein embodiment designated SHT. Source: Artificial.

SEQ ID NO: 25 is the amino acid sequence of the hybrid protein embodiment designated SLT. Source: Artificial.

5 SEQ ID NO: 26 is the amino acid sequence of the hybrid protein embodiment designated SMT. SOURCE: ARTIFICIAL.

SEQ ID NO: 27 is the amino acid sequence of the hybrid protein embodiment designated SZT Source: Artificial.

SEQ ID NO: 28 is a DNA sequence comprising the pDOJHR ORI. Source: Bifidobacterium.

10 SEQ ID NO: 29 is a DNA sequence comprising the pUC/E. coli ORI. Source: E. coli.

SEQ ID NO: 30 is a DNA sequence comprising the pB44 ORI.Source: Bifidobacterium.

SEQ ID NO: 31 is the DNA sequence encoding the TAT domain. Source: Human HIV virus.

SEQ ID NO: 32 is the DNA sequence encoding the P-beta MPG (gp41-SV40) Transduction domain. Source: SV40 virus.

15 SEQ ID NO: 33 is the DNA sequence encoding the Transportan (Galanin-mastoparan) transduction domain Source: eukaryote – species unknown,

SEQ ID NO: 34 is the DNA sequence encoding the Pep-1 (Trp-rich motif-SV40) transduction domain. Source: SV40.

SEQ ID NO: 35 GFP forward primer Source: Aequorea victoria

20 SEQ ID NO: 36 GFP Reverse Primer. Source: Aequorea victoria

SEQ ID NO: 37 Spectinomycin Resistance Forward Primer. Source: Streptomyces spectabilis.

SEQ ID NO: 38 Spectinomycin Reverse Primer: Source: Streptomyces spectabilis.

SEQ ID NO: 39 CMV Forward Primer. Source: Cytomegalovirus.

SEQ ID NO: 40 CMV Reverse Primer. Source: Cytomegalovirus.

25 SEQ ID NO: 41 MerR DNA recognition site. Source: Bifidobacterium.

SEQ ID NO: 42 LacI repressor DNA binding site. Source. E. Coli.

SEQ ID NO: 43 Synthetic zinc finger DNA recognition site. Source: Artificial.

SEQ ID NO: 44 HU DNA binding domain. Source: Bifidobacterium.

SEQ ID NO: 45 Mer R DNA binding domain. Source: Bifidobacterium.

SEQ ID NO: 46 Zn finger DNA binding domain. Source: Artificial.

SEQ ID NO: 47 Lac I DNA binding domain. Source: E. coli. E. coli.

SEQ ID NO: 48 Prokaryotic expression cassette from pFRG (FIG. 2) comprising Hu Promoter and terminator. Source: Artificial/bifidobacteria.

SEQ ID NO: 49 Eukaryotic Expression Cassette from pFRG (FIG. 2) comprising CMV promoter, Kozak sequence and TK Poly A site and terminator, flanking GFP coding sequences to be expressed . Source: Artificial.

Definition of Terms: In this disclosure, the word “comprising” is used in a non-limiting sense to mean that items following the word are included, but items not specifically mentioned are not excluded. A reference to an element by the indefinite article “a” does not exclude the possibility that more than one of the elements is present, unless the context clearly requires that there be one and only one of the elements.

In this disclosure the recitation of numerical ranges by endpoints includes all numbers subsumed within that range including all whole numbers, all integers and all fractional intermediates (e.g., 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.80, 4, and 5 etc.).

In this disclosure the singular forms a "an", and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to a composition containing "a compound" includes a mixture of two or more compounds.

In this disclosure term "or" is generally employed in its sense including "and/or" unless the content clearly dictates otherwise.

In this disclosure, unless otherwise indicated, all numbers expressing quantities or ingredients, measurement of properties and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about". Accordingly, unless indicated to the contrary or necessary in light of the context, the numerical parameters set forth in the

disclosure are approximations that can vary depending upon the desired properties sought to be obtained by those skilled in the art utilizing the teachings of the present disclosure and in light of the inaccuracies of measurement and quantification. Without limiting the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the disclosure are approximations, their numerical values set forth in the specific examples are understood broadly only to the extent that this is consistent with the validity of the disclosure and the distinction of the subject matter disclosed and claimed from the prior art.

The nucleic acid vectors, also referred to as expression vectors, disclosed herein are able to replicate in both gram negative and gram positive bacteria. In embodiments the gram negative bacteria is *E. coli*. In embodiments the gram positive bacteria is bifidobacteria. In embodiments the gram positive bacterium is *staphylococcus*. In embodiments the gram positive bacterium is *streptococcus*. In embodiments the gram positive bacterium is *lactococcus*. In embodiments the gram positive bacterium is *clostridium* or is *lactobacillus*. The full range of possible target bacterial strains will be readily understood by one skilled in the art, and a listing of strains is available at LPSN bacterio.net at http://www.bacterio.net/-alintro.html, the contents of which is hereby incorporated herein where permissible by law. In particular embodiments the bacteria are probiotic bacteria or are GRAS bacteria.

In embodiments where the target gram positive bacteria is *Bifidobacteria*, then illustrative possible strains or species of *bifidobacteria* , without limitation, include:

Bifidobacterium, Bifidobacterium actinocoloniiforme, Bifidobacterium adolescentis, Bifidobacterium angulatum, Bifidobacterium animalis, Bifidobacterium animalis subsp. *animalis, Bifidobacterium animalis* subsp. *lactis, Bifidobacterium asteroids, Bifidobacterium biavatii, Bifidobacterium bifidum, Bifidobacterium bohemicum, Bifidobacterium bombi, Bifidobacterium boum, Bifidobacterium breve, Bifidobacterium callitrichos, Bifidobacterium catenulatum, Bifidobacterium choerinum, Bifidobacterium coryneforme, Bifidobacterium vectocuniculi, Bifidobacterium denticolens, Bifidobacterium dentium, Bifidobacterium gallicum, Bifidobacterium gallinarum, Bifidobacterium globosum, Bifidobacterium indicum, Bifidobacterium infantis, Bifidobacterium inopinatum, Bifidobacterium kashiwanohense, Bifidobacterium lactis, Bifidobacterium longum, Bifidobacterium longum* subsp. *infantis, Bifidobacterium longum* subsp. *longum, Bifidobacterium longum* subsp. *suis, Bifidobacterium magnum, Bifidobacterium*

merycicum, Bifidobacterium minimum, Bifidobacterium mongoliense, Bifidobacterium pseudocatenulatum, Bifidobacterium pseudolongum, Bifidobacterium pseudolongum subsp. *globosum, Bifidobacterium pseudolongum* subsp. *pseudolongum, Bifidobacterium psychraerophilum, Bifidobacterium pullorum, Bifidobacterium reuteri, Bifidobacterium ruminantium, Bifidobacterium saeculare, Bifidobacterium saguini, Bifidobacterium scardovii, Bifidobacterium stellenboschense, Bifidobacterium stercoris, Bifidobacterium subtile, Bifidobacterium suis, Bifidobacterium thermacidophilum, Bifidobacterium thermacidophilum* subsp. *porcinum, Bifidobacterium thermacidophilum* subsp. *thermacidophilum, Bifidobacterium thermophilum, Bifidobacterium tsurumiense; Bifidobacterium longum, Bifidodobacterium bifidum, and Bifidobacterium infantis.*

Where the target bacterium is staphylococcus then the illustrative possible strains or species of staphylococcus, without limitation, include:

S. arlettae; S. agnetis; S. aureus; S. auricularis; S. capitis; S. caprae; S. carnosus; S. caseolyticus; S. chromogenes; S. cohnii; S. condimenti; S. delphini; S. devriesei; S. epidermidis; S. equorum; S. felis; S. fleurettii; S. gallinarum; S. haemolyticus; S. hominis; S. hyicus; S. intermedius; S. kloosii; S. leei; S. lentus; S. lugdunensis; S. lutrae; S. massiliensis; S. microti; S. muscae; S. nepalensis; S. pasteuri; S. pettenkoferi; S. piscifermentans; S. pseudintermedius; S. pseudolugdunensis; S. pulvereri; S. rostri; S. saccharolyticus; S. saprophyticus; S. schleiferi; S. sciuri; S. simiae; S. simulans; S. stepanovicii; S. succinus; S. vitulinus; S. warneri; and *S. xylosus.*

In this disclosure the term “vector” refers to at least one of a plasmid, bacteriophage, cosmid, artificial chromosome, or other nucleic acid vector. In embodiments the vector encodes or is suitable to generate at least one therapeutic agent or comprises at least one therapeutic sequence. Vectors suitable for microbiological applications are well known in the art, and are routinely designed and developed for particular purposes. Some non-limiting published examples of vectors that have been used to transform bacterial strains include the following plasmids: pMW211, pBAD-DEST49, pDONRP4-P1R, pENTR-PBAD, pENTR-DUAL, pENTR-term, pBR322, pDESTR4-R3, pBGS18-N9uc8, pBS24Ub, pUbNuc, pIXY154, pBR322DEST, pBR322DEST-PBAD-DUAL-term, pJIM2093, pTG2247, pMEC10, pMEC46, pMEC127, pTX, pSK360, pACYC184, pBOE93, pBR327, pDW205, pKCL11, pKK2247, pMR60, pOU82, pR2172, pSK330, pSK342, pSK355, pUHE21-2, pEHLYA2-SD. See, for example, Stritzker, et al. Intl. J. Med. Microbiol. Vol. 297, pp. 151-162 (2007); Grangette et al., Infect. Immun. vol. 72,

pp. 2731-2737 (2004), Knudsen and Karlstrom, App. and Env. Microbiol. pp. 85-92, vol. 57, no. 1 (1991), Rao et al., PNAS pp. 11193-11998, vol. 102, no. 34 (2005), each of which is incorporated herein by reference. Those skilled in the art, in light of the teachings of this disclosure, will understand that alternative embodiments of the subject matter claimed herein are possible and will understand how to combine sequences taken from known vectors in order to construct such alternative embodiments. By way of example, existing vectors from the foregoing list may be modified by inserting additional origins of replication, or additional expression cassettes (non-limiting examples of expression cassettes are presented as SEQ ID NOs 48 and 49) comprising suitable promoter and termination sequences, or additional DNA binding sequences, or coding sequences for the hybrid protein (NHP) described herein. Those skilled in the art will understand and adopt the various alternatives known in the art, and will do so using techniques well known in the art. In particular embodiments, suitable origins of replication for use in embodiments include an origin of replication comprised within pDOJHR as well as the pB44 and pUC (E. coli) origins of replication. These are presented herein as SEQ ID NOs 28, 30, 29 respectively. In some embodiments adjacent functional components of a vector are joined by linking sequences.

In embodiments the vector comprises a eukaryotic expression cassette (a non-limiting example comprising GFP coding sequences is presented as SEQ ID NO: 49) containing a marker sequence to confirm both transformation and gene expression in the target eukaryotic cell. It will be understood that in alternative embodiments a range of alternative marker proteins and sequences are possible and for example in selected embodiments and without limitation the marker sequence encodes GFP (green fluorescent protein), RFP (red fluorescent protein), CAT (chloramphenicol acetyltransferase), luciferase, GAL (beta-galactosidase), or GUS (beta-glucuronidase). Those skilled in the art will readily understand and use all such marker sequences and reporter genes using standard techniques and materials readily available in the art.

Vectors according to embodiments comprise one or more prokaryotic expression cassettes. A non-limiting example of a prokaryotic expression cassette suitable for the expression of sequences in gram positive bacteria, and in embodiments in staphylococcus and bifidobacteria is presented as SEQ ID NO: 48. In embodiments such a cassette comprises a hybrid protein and embodiments of such hybrid proteins are disclosed herein.

It will be understood that in embodiments vectors may comprise at least or only one, two, three or more eukaryotic expression cassettes, or may comprise at least or only one, two, three or more prokaryotic expression cassettes for expression in gram positive bacteria, or may comprise combinations of the foregoing.

In this disclosure the term Cell penetrating peptide ("CPP") means a protein that is able to penetrate the cell membrane of a eukaryotic cell. The term CPP includes Tat, also referred to as a trans-activator of transcription. For greater certainty, but without limitation, reference to Tat or any other CPP will be understood to mean the native sequence, and also a full range of sequence variants thereof which are suitable to carry out the desired function of the protein or protein domain in question. By way of example, a range of functional variants of the Tat protein are described in F. Salomone et al., (2010) "A novel chimeric cell-penetrating peptide with membrane-disruptive properties for efficient endosomal escape" Journal of Controlled Release 163, 293-303.

Other non-limiting examples of CPP's according to embodiments include the VP22 protein of Herpes Simplex Virus, and the protein transduction domain of the Antennapedia (Antp) protein as well as the protein transduction domains presented in Table 1, namely Tat, Rev, Antp, VP22, Pep1 and Transportan.. In embodiments the CPP domain is the domain described in Salomone F., et al, "A novel chimeric cell-penetrating peptide with membrane disruptive properties for endosomal escape." J Control. Release. 2012 Oct 2 (epub ahead of print). Table 1 shows a non-limiting selection of exemplary CPP domains.

TABLE 1: Sequences of exemplary CPP (transduction) domains

Tat	SEQ ID NO: 11	YGRKKRRQRRR
Antp	SEQ ID NO: 12	RQIKIWFQNRRMKWKK
Rev	SEQ ID NO: 13	TRQARRNRRRRWRERQF
VP22	SEQ ID NO: 14	NAKTRRHERRRKLAIER
P-beta MPG (gp41-SV40)	SEQ ID NO: 15	GALFLGFLGAAGSTMGAWSQPKKKRKV-cys
Transportan (Galanin-mastoparan)	SEQ ID NO: 16	GWTLNSAGYLLGKINLKALAALAKKIL
Pep-1 (Trp-rich motif-SV40)	SEQ ID NO: 17	KETWWETWWTEWSQPKKKRRV

In this disclosure the term “DNA binding domain” means a protein sequence able to reversibly but tightly or with high affinity bind specifically to a suitable DNA sequence. In embodiments a DNA binding sequence may be a Zinc finger binding domain, and while a wide range of suitable domains and their complementary DNA binding sequences will be readily identified by those skilled in the art, a number of illustrative examples are disclosed in US Patent No. 6007988, issued on December 28, 1999. In particular embodiments hereof, the DNA binding motif or domain is or is derived from, Mer R, Zinc finger, or Histone like DNA binding protein or is or is derived from the HU protein or is or is derived from a homeobox DNA binding protein. It will be understood that the HU protein is generally considered a homeobox-like protein.

While many types of DNA binding domains will be readily identified by those skilled in the art using available databases, screening methodologies and well known techniques, non-limiting examples of suitable DNA binding domains for use in alternative embodiments can be derived from a wide range of DNA binding proteins. In embodiments suitable DNA binding domains may be of any general type, including but not limited to helix-turn-helix, Zinc finger, leucine zipper, winged helix, winged helix turn helix, helix loop helix, HMG box, Wor 3 and RNA guided binding domains. Illustrative examples of DNA binding proteins whose DNA binding domains may be utilized in embodiments include histones, histone like proteins, transcription promoters, transcription repressors, transcriptional regulators, which may be drawn from a wide range of alternate sources and operons.

In this disclosure the terms "polypeptide", "peptide", “oligopeptide” and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residues is an artificial chemical analogue of a corresponding naturally occurring amino acid, or is a completely artificial amino acid with no obvious natural analogue as well as to naturally occurring amino acid polymers.

In embodiments the eukaryotic expression cassettes comprised in the vectors comprise suitable Kozak sequences, and the possible variations thereon and positioning thereof will be readily understood by those skilled in the art. One non limiting example is presented as SEQ ID NO: 49.

A peptide or peptide fragment is "derived from" a parent peptide or polypeptide if it has an amino acid sequence that is homologous to the amino acid sequence of, or is a conserved fragment from, the parent peptide or polypeptide. It will be understood that such sequences will,

in alternative embodiments, comprise natural amino acids or will comprise artificially created amino acids. All of the foregoing will be readily identified by those skilled in the art.

In embodiments vectors have sequences which differ from the disclosed examples. It will be understood by those skilled in the art that a range of sequence variations are possible that do not affect or do not prevent the suitability of the vector for the purposes disclosed herein. By way of example and not limitation those skilled in the art will recognise that certain DNA sequences can be varied without materially affecting the function of the vector and others cannot. Again by way of illustration and not limitation, those skilled in the art will recognise and adopt sequence modifications known to enhance the function of a selected sequence, and will reject sequence modifications known to diminish such function. With particular reference to protein coding sequences those skilled in the art will recognise variable and conserved regions and will recognise mutations likely to change the structure and/or function of the relevant protein or polypeptide and those unlikely to do so. In general it will be understood that in particular variant embodiments the nucleic acid sequence of a vector will be 100% identical to one of the examples disclosed, or will be at least about, or about, or less than about 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 95%, 84%, 83%, 82%, 81%, or 80% identical to one of the examples disclosed and that in embodiments such sequence identity will extend over all or only part of the length of the vector. It will be understood that in embodiments vectors will comprise alternative origins of replication, promoters, polyadenylation sites and the like. In embodiments vectors will comprise sequences inserted in the expression cassettes, and in embodiments vectors will have no insertions in the expression cassettes.

In embodiments the hybrid protein or polypeptide sequence encoded by a sequence comprised in an expression cassette according to an embodiment is 100% identical to one of the examples disclosed, or is at least about, or about, or less than about 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 95%, 84%, 83%, 82%, 81%, or 80% identical to one of the examples disclosed as SEQ ID NOs 24, 25, 26, 27 and such sequence identity extends over all or only part of the length of the protein or polypeptide.

In embodiments homologies and sequence identities extend over all or only a part of the sequence or sequences of interest. In embodiments homologies and sequence identities are limited to particular functional or sequence domains. In embodiments, homologies and sequence identities continuous and in embodiments are separated by regions of lower sequence identity or homology.

Construction of vectors: In embodiments vectors and sequences set out herein are synthesized *de novo* using known techniques and commercially available DNA synthesis services. Standard techniques for the construction of the vectors of the present invention are well-known to those of ordinary skill in the art and can be found in such references as Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York, (1989). A variety of strategies are available for ligating fragments of DNA, the choice of which depends on the nature of the termini of the DNA fragments and which choices can be readily made by the skilled artisan. In embodiments the sequences of vectors are determined using suitable DNA, RNA and protein sequence manipulation software, and the desired sequence is synthesized using suitable synthesis methods, a wide variety of which are readily available and will be immediately understood and implemented by those skilled in the art.

As described herein, an aspect of the present disclosure concerns isolated nucleic acids and methods of use of isolated nucleic acids. Plasmid Preparations: Plasmid preparations and replication means are well known in the art. See for example, U.S. Pat. Nos. 4,273,875 and 4,567,146 incorporated herein in their entirety. Some embodiments of the present invention include providing a portion of genetic material of a target microorganism and inserting the portion of genetic material of a target microorganism into a plasmid for use as an internal control plasmid.

Nucleic acids used as a template for amplification are isolated from cells according to standard methodologies. (Sambrook et al., 1989) The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary cDNA. Pairs of primers that selectively hybridize to nucleic acids corresponding to specific sequences are contacted with the isolated nucleic acid under conditions that permit selective hybridization. Once hybridized, the nucleic acid primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

Next, the amplification product is detected. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintilography of incorporated radiolabel or fluorescent label or even via a system using electrical or thermal impulse signals (Affymax technology; Bellus, 1994).

Primers: The term primer, as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty base pairs in length, but longer sequences may be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred. Specific primers used to amplify portions of vectors and nucleotide sequences according to embodiments are presented as SEQ ID NOs 35-40. Those skilled in the art will readily select alternative suitable primers for particular requirements.

Template Dependent Amplification Methods: A number of template dependent processes are available to amplify the sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR) which is described in detail in U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis et al., 1990, each of which is incorporated herein by reference in its entirety.

A reverse transcriptase PCR amplification procedure may be performed in order to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook et al., 1989. Alternative methods for reverse transcription utilize thermostable DNA polymerases. These methods are described in WO 90/07641 filed Dec. 21, 1990. Polymerase chain reaction methodologies are well known in the art. Other amplification methods are known in the art besides PCR such as LCR (ligase chain reaction), disclosed in European Application o. 320 308, incorporated herein by reference in its entirety.

In another embodiment, Qbeta Replicase, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA which has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence which may then be detected.

Nucleic Acid Synthesis: Those skilled in the art will readily recognize a range of methods and apparatuses for synthesizing desired nucleic acid sequences. By way of example and not of limitation in a series of embodiments plasmids were synthesized in silico by GeneArt®, Life Technologies™

While the scope of the of methods for making embodiments includes any suitable methods (for example, Polymerase Chain Reaction, i.e., PCR, and nucleic acid sequence based amplification, i.e., NASBA) for amplifying at least a portion of the microorganism's genetic

material, for one example, the present invention describes embodiments in reference to PCR technique.

Amplification of a genetic material, e.g., DNA, is well known in the art. See, for example, U.S. Pat. Nos. 4,683,202, and 4,994,370, which are incorporated herein by reference in their entirety. By knowing the nucleotide sequences of desired genetic material or target nucleic acid sequence, specific primer sequences can be designed In one embodiment of the present invention, the primer is about, but not limited to 5 to 50 oligonucleotides long, or about 10 to 40 oligonucleotides long or more about 10 to 30 oligonucleotides long. Suitable primer sequences can be readily synthesized by one skilled in the art or are readily available from third party providers such as BRL (New England Biolabs™), etc. Other reagents, such as DNA polymerases and nucleotides, that are necessary for a nucleic acid sequence amplification such as PCR are also commercially available.

Separation Methods: Following amplification, it may be desirable to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods. See Sambrook et al., 1989.

Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography (Freifelder, 1982).

Identification Methods: Amplification products must be visualized in order to confirm amplification of the desired sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products may then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled, nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In

another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, where the other member of the binding pair carries a detectable moiety.

In one embodiment, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art and may be found in many standard books on molecular protocols. See Sambrook et al., 1989. Briefly, amplification products are separated by gel electrophoresis. The gel is then contacted with a membrane, such as nitrocellulose, permitting transfer of the nucleic acid and non-covalent binding. Subsequently, the membrane is incubated with a chromophore-conjugated probe that is capable of hybridizing with a target amplification product. Detection is by exposure of the membrane to x-ray film or ion-emitting detection devices.

In general, prokaryotes used for cloning DNA sequences in constructing the vectors useful in the invention include for example, any gram negative bacteria such as E. coli, *E. coli* strain K12 and *bifidobacteria* and *staphylococcus* Other microbial strains which may be used include *P. aeruginosa* strain PAO1, and *E. coli* B strain and *bifidobacteria*. These examples are illustrative rather than limiting. In particular embodiments steps in the construction of vectors may include cloning and propagation in suitable *E. coli* strains or in suitable *bifidobacterium* strains.

In general, plasmid vectors containing promoters and control sequences which are derived from species compatible with the host cell are used with these hosts. The vector ordinarily carries a replication site as well as one or more marker sequences which are capable of providing phenotypic selection in transformed cells. For example, a PBBR1 replicon region which is useful in many Gram negative bacterial strains or any other replicon region that is of use in a broad range of Gram negative host bacteria can be used in the present invention.

The term "recombinant polypeptide", “recombinant protein: or “fusion protein” or “hybrid protein” or like terms is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide. In particular embodiments, fusion proteins contain joining sequences to join protein domains that are not normally associated. In broad concept, in embodiments, fusion proteins comprise at least one sequence selected from the group consisting of: a DNA binding domain, a secretion signal sequence, and a trans-activator of transcription that is functional in Eukaryotic cells. In further embodiments fusion proteins comprise a CPP domain and in embodiments the CPP domain comprises a TAT domain. The

hybrid proteins comprising signal sequence, transduction domain and DNA binding domain, are also referred to herein as simply "hybrid proteins" or as "NHPs" depending on the context. It will be understood that the different domains comprised in hybrid proteins according to embodiments may be present in any arrangement and any combination and any numbers of copies consistent with their function. Thus in embodiments signal sequences may be internal or may be terminal signal sequences and there may be multiple copies of one or more of each of the domains. In embodiments vectors may comprise one, two, three, four or more DNA sequences suitable for binding by the one or more DNA binding domains comprised in a hybrid protein according to embodiments.

In embodiments vectors comprise a prokaryotic expression cassette for expressing an NHP in the gram positive bacterium and a eukaryotic expression cassette for expressing a candidate gene or sequence in a eukaryotic cell. .It will be understood that in alternative embodiments the prokaryotic expression cassette will contain other proteins desired to be expressed in the target gram positive bacterium, instead of one of the hybrid proteins disclosed herein. Thus in embodiments, vectors may be used to express a sequence or gene of interest in a prokaryotic cell.

In this disclosure the terms SHT, SZT, SLT, SMT are acronyms indicating currently preferred forms of the NHP. In the foregoing acronym descriptors Z indicates the presence of at least one Zinc finger DNA binding domain an example of the DNA sequence coding for which is presented as SEQ ID NO: 6 and the amino acid sequence as SEQ ID NO: 46 , H means the presence of at least one HU DNA binding domain an example of the DNA sequence coding for which is presented as SEQ ID NO: 44 and the amino acid sequence at SEQ ID NO: 44, T means a CPP or protein transduction domain examples of which are presented as SEQ ID NOs 11 through 17. S means a suitable signal sequence examples of which are presented as SEQ ID NOs 18 through 23. M means the presence of at least one Mer R DNA binding domain an example of the DNA sequence coding for which is presented as SEQ ID NO: 5 and the protein sequence of which is presented at SEQ ID NO: 45. L represents a Lac-I DNA binding domain which is presented herein as SEQ ID NO: 47, it being understood that the inventors have found that hybrid proteins comprising the Lac-I DNA binding domain are not effective for the purposes hereof and are not embodiments of the subject matter claimed herein.

In embodiments presented the signal sequence is an alpha-L-arabinosidase signal sequence SEQ ID NOs 22, 23 but in alternative embodiments is the alpha amylase signal sequence SEQ

ID NOs 18, 19 or a truncated version thereof SEQ. ID NOS. 20, 21, are non-limiting alternatives. In embodiments the transduction domain T is Tat SEQ ID NO: 11, but a number of non-limiting possible alternatives are presented herein. Sequences of exemplary SHT, SMT, SZT hybrid proteins are presented as SEQ ID NOs 24, 26 and 27.

It will be understood that variant signal domains, CPP domains and DNA binding domains may be comprised in the hybrid proteins. Thus by way of example and not limitation, the S or signal sequence domain may be any suitable signal sequence, non-limiting examples being the alpha arabinosidase signal sequence, the alpha amylase signal sequence, and truncated versions of these and other signal sequences. The foregoing sequences, or their corresponding DNA sequence, are presented as SEQ ID NOs 18 through 23.

Possible CPP domains (identified as "T" in the abbreviations for the hybrid proteins according to embodiments) include Tat, Antp. Rev, VP22, PbetaMPG (gp41-SV40), Transportan (galanan mastoparan) and Pep1 (TRP rich motif SV40), or the biologically effective portions thereof. Amino acid sequences of the foregoing are presented as SEQ ID NOs 11-17.

It will be understood that in embodiments the foregoing hybrid protein constructions will be comprised in the prokaryotic gene expression cassette of the vector and the foregoing prokaryotic expression cassette will be designed to be functional in a desired bacterial strain. In particular embodiments such strain is a gram positive bacterium and in embodiments is bifidobacterium. In embodiments it will be staphylococcus. It will likewise be understood that in embodiments the desired DNA for transformation into a target cell comprises a mammalian gene expression cassette for expression of a candidate sequence in the target cell. In embodiments the mammalian cell is a human cell.

The term "expression cassette" is used herein to describe a DNA sequence context comprising one or more locations into which a selected sequence may be inserted so as to be expressible as RNA. In general an expression cassette will comprise a promoter, transcription start site, transcription termination site. A sequence inserted at an appropriate location, between the transcription start and termination sites, will then be expressed when the cassette is introduced into a suitable host cell. In embodiments the choices of flanking sequences will be chosen to function in a chosen cell. Those skilled in the art will thus insert suitable sequences to be expressed in appropriate expression cassettes so that sequences desired to be expressed in a prokaryotic cell will be in a suitable sequence context and those desired to be expressed in eukaryotic cells will likewise be in a suitable sequence context. Those skilled in the art will

readily adjust the sequences, structure and other aspects of the cassette to suit particular purposes or to function in particular bacterial strains. In one series of embodiments the coding sequences for a hybrid protein according to embodiments is bounded by a HU promoter and transcription initiation site and a HU transcription termination site. In another embodiment (pFRG1.5, FIG. 1, SEQ ID NO: 2) the promoter is a ribosomal RNA promoter found in bifidobacteria, and the terminator is a ribosomal RNA terminator found in bifidobacteria. In embodiments the eukaryotic expression cassette comprises a suitable Kozak sequence.

In this disclosure the term "secretion signal", "secretion sequence", "secretion signal sequence", "signal sequence" and the like, refer to a protein motif that is effective to cause secretion of the protein across a cell membrane. In embodiments a secretion sequence is, or is derived from the alpha-L-arabinosidase signal sequence or is an alpha-amylase signal sequence or is a truncated alpha amylase signal sequence. Protein sequences for the foregoing are presented as SEQ ID NOs 23, 19 and 21 respectively and the encoding DNA sequences are presented as SEQ ID NOs 22, 18 and 20 respectively. Those skilled in the art will readily identify and implement suitable signal sequences which are useable in alternative embodiments. One listing of possible signal sequences is to be found in the signal sequence database at <www.signalpeptide.de> and in other resources such as SPdb – Signal Peptide Resource at <http://proline.bic.nus.edu.sg>.

While thousands of suitable signal sequences will be identified by those skilled in the art, using available databases, screening methodologies and well known techniques, non-limiting examples of sources for suitable signal sequences for use in alternative embodiments can optionally be derived from a range of secreted enzymes and other proteins, examples including carbohydrases such as amylases, sucrases, galactosidases, monoshaccharide transferrases, lipases, phospholipases, reductases, oxidases, peptidases, transferases, methylases, ethylases, cellulases, ligninases, secreted signalling proteins, toxins and all manner of other secreted proteins.

Construction of suitable vectors containing the desired coding and control sequences can be achieved employing standard ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to form the plasmids required. In embodiments the vectors are synthesized de novo using suitable nucleic acid synthesis procedures as explained elsewhere herein. A range of alternative promoters, polyadenylation

signals, DNA binding domains, cell penetrating peptide domains, signal sequences and the like will be readily identified by those skilled in the art using conventional methods and resources.

For analysis to confirm correct sequences in plasmids constructed, the ligation mixtures may be used to transform a bacteria strain such as *E. coli* K12 and successful transformants selected by antibiotic resistance using suitable selection markers and in embodiments the selection markers are antibiotics such as tetracycline, ampicillin spectinomycin, penicillin, kanamycin, gentamycin, zeomycin, methicillin, hygromycin B and others, all of which will be immediately recognized and used by those skilled in the art. Plasmids from the transformants are prepared, analyzed by restriction and/or sequenced. In alternative embodiments bacteria may be selected using suitable auxotrophic selection markers, non-limiting examples of which include the *LEU2* and *URA3* selection markers whose nature and use are well understood by those skilled in the art. In particular embodiments the selection marker confers resistance antibiotics effective against both gram positive and gram negative bacteria and in embodiments this is spectinomycin, tetracycline or chloramphenicol or erythromycin. In alternative embodiments the vector comprises multiple selection markers to permit selection using different antibiotics in gram positive and gram negative bacteria.

Host cells can be transformed with nucleic acid vectors of this invention and cultured in conventional nutrient media modified as is appropriate for inducing promoters, selecting transformants or amplifying genes. The culture conditions, such as temperature, pH and the like, will be apparent to the ordinarily skilled artisan.

"Transformation" refers to the taking up of vector or of a desired DNA sequence by a host cell whether or not any coding sequences are in fact expressed. Numerous methods are known to the ordinarily skilled artisan, for example, Ca salts and electroporation. Successful transformation is generally recognized when any indication of the operation of the vector or stable propagation of the introduced DNA occurs within the host cell.

It will be understood that the vectors and hybrid proteins disclosed herein are of particular value where standard or commonly used transformation procedures are not reliable.

As used interchangeably herein, the terms "nucleic acid molecule(s)", "oligonucleotide(s)", and "polynucleotide(s)" and “nucleic acids” and the like include RNA or DNA (either single or double stranded, coding, complementary or antisense), or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form (although each of the above species may be

particularly specified), as is consistent or necessary in context. The term "nucleotide" is used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide.

The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another by virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, 1995, which disclosure is hereby incorporated by reference in its entirety).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G. "Complement" is used herein as a synonym from "complementary polynucleotide", "complementary nucleic acid" and "complementary nucleotide sequence". These terms are applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind. Unless otherwise stated, all complementary polynucleotides are fully complementary on the whole length of the considered polynucleotide.

Digestion of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be

known to the ordinarily skilled artisan. For analytical purposes, typically 1 μg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Recovery or isolation of a given fragment of DNA from a restriction digest means separation of the digest on polyacrylamide or agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. This procedure is known generally (Lawn, R. et al., Nucleic Acids Res. 9: 6103 6114 [1981], and Goeddel, D. et al., Nucleic Acids Res. 8: 4057 [1980]).

Dephosphorylation refers to the removal of the terminal 5' phosphates by treatment with bacterial alkaline phosphatase (BAP). This procedure prevents the two restriction cleaved ends of a DNA fragment from "circularizing" or forming a closed loop that would impede insertion of another DNA fragment at the restriction site. Procedures and reagents for dephosphorylation are conventional (Maniatis, T. et al., Molecular Cloning, 133 134 Cold Spring Harbor, [1982]). Reactions using BAP are carried out in 50 mM Tris at 68°C. to suppress the activity of any exonucleases which may be present in the enzyme preparations. Reactions are run for 1 hour. Following the reaction the DNA fragment is gel purified.

Ligation refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T. et al., Id. at 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

Filling or blunting refers to the procedures by which the single stranded end in the cohesive terminus of a restriction enzyme-cleaved nucleic acid is converted to a double strand. This eliminates the cohesive terminus and forms a blunt end. This process is a versatile tool for converting a restriction cut end that may be cohesive with the ends created by only one or a few other restriction enzymes into a terminus compatible with any blunt-cutting restriction endonuclease or other filled cohesive terminus. In one embodiment, blunting is accomplished by

incubating around 2 to 20 μg of the target DNA in 10 mM $MgCl_2$, 1 mM dithiothreitol, 50 mM NaCl, 10 mM Tris (pH 7.5) buffer at about 37°C. in the presence of 8 units of the Klenow fragment of DNA polymerase I and 250 μM of each of the four deoxynucleoside triphosphates. The incubation generally is terminated after 30 min. phenol and chloroform extraction and ethanol precipitation.

The terms "polypeptide" and "protein", used interchangeably herein, refer to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude chemical or post-expression modifications of the polypeptides of the invention, although chemical or post-expression modifications of these polypeptides may be included excluded as specific embodiments. Therefore, for example, modifications to polypeptides that include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Further, polypeptides with these modifications may be specified as individual species to be included or excluded from the present invention. The natural or other chemical modifications, such as those listed in examples above can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance Creighton (1993); Seifter et al., (1990); Rattan et al., (1992)). Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems, etc.), polypeptides with substituted linkages, as

well as other modifications known in the art, both naturally occurring and non-naturally occurring.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A sequence which is "operably linked" to a regulatory sequence such as a promoter means that said regulatory element is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the nucleic acid of interest. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence.

In this disclosure the term “shuttle vector” means a DNA vector which is able to replicate in both gram positive and gram negative bacteria and comprises a eukaryotic expression cassette suitable to express a sequence or gene of interest when introduced to a eukaryotic cell and a prokaryotic expression cassette able to express a gene of interest in a eukaryotic cell. In embodiments a first and second strains of host bacteria are from different genera, in embodiments they are from different species, and in embodiments they are from different subspecies. In embodiments the gram negative bacteria is *an E. coli* and in embodiments the gram positive bacteria is *lactococcus*, *lactobacillus*, *bifidobacterium*, or *staphylococcus*. It will be understood that a wide variety of combinations of first and second strains of host bacteria are possible in alternative embodiments and the foregoing exemplary combination of *E. coli* with other strains is in no way limiting. In embodiments shuttle vectors are plasmids. Thus the nucleic acid vectors disclosed herein are shuttle vectors able to replicate in both gram positive and gram negative bacteria of suitable types.

A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a gene. The phrase "under transcriptional control" means that the promoter is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the gene. The particular promoter employed to control the expression of a nucleic acid sequence of interest is not believed to be important, so long as it is capable of directing the expression of the nucleic acid in the targeted cell.

In embodiments the target eukaryotic cell is a mammalian cell. In embodiments the mammalian cell is a human cell and in embodiments is a cancer cell. In embodiments the cancer cell is a lung cancer cell, a colon (colorectal) cancer cell, a kidney cancer cell, or an ovarian cancer cell. In particular embodiments the cell is HEK-29 (human embryonic kidney); HT29, CaCo2 (both

human adenocarcinoma); HeLa, LL2 (lung carcinoma). In embodiments the cells are cultured cells.

Where a cDNA insert is employed, typically one will typically include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed. Also contemplated as an element of the expression construct is a terminator. These elements can serve to enhance message levels and to minimize read through from the construct into other sequences.

Kits: In some embodiments, there are disclosed kits comprising the vectors disclosed herein. In embodiments the kits comprise a quantity of vector DNA. Embodiments of exemplary nucleic acid vectors and their sequences are presented as FIG. 1 and 2 and SEQ ID NOs 1 and 2. In embodiments kits comprise a quantity of NHP. In embodiments the kits comprise instructions.

The container means of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which the vector, cells and/or primers may be placed, and preferably, suitably aliquoted. Where an additional component is provided, the kit will also generally contain additional containers into which this component may be placed. The kits of the present invention will also typically include a means for containing the probes, primers, and any other reagent containers in close confinement for commercial sale. In embodiments a kit will include injection or blow-molded plastic containers into which the desired vials are retained and in embodiments a kit will include instructions regarding the use of the materials comprised in the kit.

In this disclosure “gene of interest”, “sequence of interest”, “candidate gene” or “cargo gene” and like terms, refer to a sequence or gene that it is desired to transform into a target cell or that it is desired to express in a target cell. In embodiments hereof such sequences and genes are incorporated into vectors capable of expression of such sequence in particular target cells.

In this disclosure the term “target cell” means a cell into which it is desired to introduce a candidate or cargo gene or sequence. In embodiments a target cell is a eukaryotic cell and in embodiments is a mammalian cell. In embodiments a target cell is a prokaryotic cell.

In embodiments, vectors contain sequences necessary for efficient transcription and translation of specific genes or sequences encoding specific mRNA or siRNA sequences, in a target probiotic cell and may thus comprise transcription initiation and termination sites, enhancers and

the like. Similarly any expressed RNA may comprise suitable translation start sites, ribosome binding sites, and the like. All of which will be readily identified by those skilled in the art.

More generally the term "expression cassette" is used herein to describe a DNA sequence context comprising one or more locations into which a selected sequence may be inserted so as to be expressible when present in a suitable cell type. In general an expression cassette will comprise a promoter, transcription start site, transcription termination site and other necessary or desirable sequences. In embodiments the expression cassette comprises a multiple cloning site suitable to permit the convenient insertion thereinto of a DNA sequence having compatible ends, and is able to be expressed as a translatable RNA by suitable cell types. In embodiments, rather than having a multiple cloning site, each expression cassette has restriction cut sites flaking 5' and 3' ends of promoter, gene of interest and terminator. It will be understood that in order for a protein to be expressed the insertion of the DNA sequence must be in the correct reading frame. A sequence inserted at an appropriate location, between the transcription start and termination sites, will then be expressed when the cassette is introduced into a suitable host cell and in embodiments such host cell is bifidobacterial cell. It will be understood that where a chosen nucleotide sequence is said to be inserted into an expression cassette, or where it is said that an expression cassette comprises or includes a chosen nucleotide sequence, then such chosen nucleotide sequence will be inserted into such expression cassette in a form suitable for transcription and/or translation to generate a biologically active form of any protein or oligopeptide encoded thereby or will be expressible in the form of a suitable RNA species. Those skilled in the art will readily adjust the sequences, structure and other aspects of the cassette to suit particular purposes or to function in particular bacterial strains.

For clarity, this disclosure refers to both first and second expression cassettes non-limiting examples of which are presented as SEQ ID NOs 48, 49. In embodiments the first expression cassette is a prokaryotic expression cassette a non-limiting example being presented as SEQ ID NO: 48 and serves to express a desired fusion protein according to embodiments and the second expression cassette is a eukaryotic expression cassette and in embodiments serves to express a candidate DNA in a target cell. Thus in embodiments fusion proteins are encoded by a prokaryotic gene expression cassette and a cargo gene or sequence of interest is comprised in a eukaryotic or mammalian gene expression cassette.

In embodiments the construction of suitable vectors containing the desired coding and control sequences is achieved using standard ligation techniques. Isolated plasmids or DNA fragments

are cleaved, tailored, and religated in the form desired to form the plasmids required. In alternative embodiments vectors, inserts, plasmids and any other nucleic acid sequences of interest, are synthesized *de novo* using standard nucleic acid synthesis techniques. In series of embodiments plasmids were synthesized in silico by GeneArt®, Life Technologies™, (Gene Art/Life Technologies™, Im Gewerbepark B35, Regensburg, 93059 Germany). Examples of other commercial nucleic acid synthesis providers are: DNA2.0 1140 O'Brian Drive Suite A, Menlo Park CA, 94025 whose website is to be found at www.dna20.com; and Genewiz, 115 Corporate Boulevard, South Plainfield, NJ 07080 with a website at www.genewiz.com.

Without limitation, the direct synthesis of desired sequences of nucleic acids and amino acids will be readily achieved by those skilled in the art using a range of known techniques. Exemplary references describing relevant synthetic methods are described in the following publications, the content of which is incorporated herein in its entirety to the full extent permissible by law: Khorana HG, Agarwal KL, Büchi H et al. (December 1972). "Studies on polynucleotides. 103. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast". *J. Mol. Biol.*72 (2): 209–217; Itakura K, Hirose T, Crea R et al. (December 1977). "Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin". *Science* 198 (4321): 1056–1063; Edge MD, Green AR, Heathcliffe GR et al. (August 1981). "Total synthesis of a human leukocyte interferon gene". *Nature* 292 (5825): 756–62.; "Difficult to Express Proteins". *Sixth Annual PEGS Summit.* Cambridge Healthtech Institute. 2010.; Liszewski, Kathy (1 May 2010). "New Tools Facilitate Protein Expression".*Genetic Engineering & Biotechnology News.* Bioprocessing 30 (9) (Mary Ann Liebert). pp. 1, 40–41.; Welch M, Govindarajan M, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C (2009). Kudla, Grzegorz, ed."Design Parameters to Control Synthetic Gene Expression in *Escherichia coli*".*PLoS ONE* 4 (9): e7002; "Protein Expression". DNA2.0. Retrieved 11 May 2010.; Fuhrmann M, Oertel W, Hegemann P (August 1999). "A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*". *Plant J.* 19 (3): 353–61.; Mandecki W, Bolling TJ (August 1988). "FokI method of gene synthesis". *Gene* 68(1): 101–7. ; Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL (October 1995)."Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides". *Gene* 164 (1): 49–53.; Gao X, Yo P, Keith A, Ragan TJ, Harris TK (November 2003). "Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences". *Nucleic Acids Res.* 31(22): e143.; Young L, Dong Q (2004). "Two-step total gene synthesis

method". *Nucleic Acids Res.* 32 (7): e59; Hillson NH, Rosengarten RD, Keasling JD (2012). "j5 DNA Assembly Design Automation Software". *ACS Synthetic Biology* 1 (1): 14–21; Hoover DM, Lubkowski J (May 2002)."DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis". *Nucleic Acids Res.* 30(10): e43.; Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S (2006). "Gene Designer: a synthetic biology tool for constructing artificial DNA segments". *BMC Bioinformatics* 7: 285.; Tian J, Gong H, Sheng N et al. (December 2004). "Accurate multiplex gene synthesis from programmable DNA microchips". *Nature* 432 (7020): 1050–4.

Similarly those skilled in the art will immediately recognise and use a range of suitable software applications and online resources for the prediction and design of desired nucleotide and protein sequences including hybrid sequences, plasmid and other vector sequences, coding and other sequences.

By way of example and not limitation, a variety of online DNA databases referred to elsewhere herein contain sequences of a variety of signal peptides, DNA binding domains and their recognition sequences, selection markers, resistance genes, auxotrophic mutants, promoters, and origins of replication. Thus for example:

One listing of possible signal sequences is to be found in the signal sequence database at <http://www.signalpeptide.de/index.php?m=listspdbat> and further information is to be found at the main website at http://www.signalpeptide.de/.

Possible transduction sequences are described herein and these and other similar sequences will be readily identified using online resources.

A range of sequence analysis and manipulation software is readily available and used by those skilled in the art. By way of example and not limitation, suitable software for the manipulation and analysis of nucleotide and/or protein sequences includes SNAPGENE, pDRAW32, DNASTAR, BLAST, CS-BLAST, FASTA, MB-DNA Analysis software, DNADynamo, Plasma DNA, Sequencher; suitable motif prediction and analysis software includes but is not limited to FMM, PMS, eMOTIF, PHI-Blast, Phyloscan, and an exemplary library of protein motifs is I-Sites. Links to all of the foregoing are to be found on the Wikipedia web page at http://en.wikipedia.org/wiki/List_of_sequence_alignment_software which was accessed on January 19, 2015.

More generally extensive databases comprising DNA, RNA, and protein sequences include but are not limited to GenBank, RefSeq, TPA, PDB and the NCBI database at http://www.ncbi.nlm.nih.gov. These and other suitable tools and resources will be readily identified and used by those skilled in the art.

Description of Embodiments

Embodiments of the invention are hereafter described with general reference to FIGS.1 through 10 and SEQ ID NOs 1 through 49 all of which form a part of and are incorporated in this disclosure.

First Embodiment

In a first general aspect of the embodiment there is disclosed a nucleic acid vector comprising: a first origin of replication for replication in gram-positive bacteria and second origin of replication for replication in gram-negative bacteria, at least one selection markers for selection in both gram-positive and gram-negative bacteria; a first gene expression cassette functional in gram positive bacteria and a second gene expression cassette functional in a mammalian cell an in embodiments this is a human cell.

In embodiments the vector comprises a single selection marker effective in both gram positive and gram negative bacteria.

A first illustrative embodiment of a vector according to the first embodiment is shown in FIG.1 and the corresponding vector sequence is presented as SEQ ID NO: 1. The design of suitable expression cassettes will be readily understood by one skilled in the art. In the embodiment pBRA2.0 SHT the first expression cassette for expression in a gram positive bacterium comprises an HU promoter and terminator flanking the NHP encoding sequences as will be seen in FIG. 1. In the embodiment pBRA2.0 SHT the second expression cassette for expression in a eukaryotic cell, which in embodiments is a mammalian cell and in embodiments is a human cell, comprises the CMV (cytomegalovirus) promoter, a KOZAK sequence and a Thymidine Kinase polyadenylation/termination sequence, the foregoing flanking an insertion sequence for the insertion of a cargo sequence. This will be seen in SEQ ID NO: 1 wherein a GFP protein is inserted into the second expression cassette.

In particular variants of the first embodiment, a first origin of replication is functional in *E. coli* and second origin of replication is functional in at least one of *Staphylococcus* and

Bifidobacteria. In further variants a single bifunctional origin of replication is functional in E. coli and in at least one of *Staphylococcus* and *Bifidobacteria*.

In the embodiment pBRA2.0 SHT the vector comprises a pUC origin of replication for replication in *E. coli* and a pB44 origin of replication for replication in *Bifidobacterium*.

5 It will be understood that suitable origins of replication for use in a gram negative bacterium, namely *E. coli*, include the pUC origin of replication presented as SEQ ID NO: 29 and a wide range of other gram negative origins, which will be readily identified and adopted by those skilled in the art.

Suitable origins of replication for *Bifidobacterium* and other gram positive strains of bacteria
10 include PB44 ORI presented as SEQ ID NO: 30, and the pDOJHR ORI presented as SEQ ID NO: 28, within which the origin is comprised.

DNA-BS denotes a nucleotide binding site (SEQ.ID. NO. 43) for a Zinc finger DNA binding protein (SEQ ID NO: 46).

It will therefore be understood that in alternative embodiments the first gram negative origin of
15 replication will comprise a sequence having at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% sequence identity over at least 20, 40, 60, 80 or more contiguous nucleotides to the pUC (E. coli) ORI presented as SEQ ID NO: 29.

It will likewise be understood that in alternative embodiments the second or gram positive origin of replication will comprise a sequence having at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or
20 100% sequence identity over at 20, 40, 60, 80 or more contiguous nucleotides to the pB44 or pDOJHR origins of replication whose sequences are presented as SEQ ID NOs 30 and 28 respectively.

In embodiments the selection marker is effective in both gram positive and gram negative bacteria. In embodiments the selection marker is resistance to spectinomycin. In embodiments
25 the selection marker comprised in the vector is resistance to tetracycline. In embodiments the selection marker comprised in the vector is resistance to chloramphenicol.

In alternative embodiments the vector comprises separate selection markers for gram positive and gram negative bacteria. Accordingly, in embodiments the vector comprises a selection marker or a combination of selection markers selected from resistance to ampicillin, penicillin,
30 tetracycline, chloramphenicol, streptomycin, quinolone, fluoroquinolone, gentamycin, neomycin,

kanamycin, spectinomycin. Those skilled in the art will readily identify additional selection markers and the gene sequences responsible therefore. Listings of possible selection markers that can be adopted in variant embodiments are to be found in Stryer, Sambrook et al. 1989, Maniatis the contents of which are all incorporated herein by reference to wherever permissible by law.

In the illustrated embodiments the first or gram negative origin of replication is functional in *E. coli*. In the illustrated embodiments the second or gram positive origin of replication is functional in at least one of *bifidobacterium* and *staphylococcus*, In alternative embodiments the origin is functional in a bacterium selected from the group consisting of *bifidobacteria*, *lactococcus*, and *clostridium*, *staphylococcus* and *streptococcus*.

In the first embodiment the first gene expression cassette encodes a hybrid protein comprising at least one DNA binding domain, at least one cell penetrating peptide (CPP) domain, and at least one secretion signal sequence. Alternative embodiments of the NHP are possible and examples are disclosed below.

In embodiments the nucleic acid vectors according to embodiments further comprise at least one DNA motif that binds to the at least one DNA binding domain of said hybrid protein. It will be understood that in embodiments the said DNA binding domain is sequence specific and that in other embodiments, such as embodiments wherein the NHP comprises an HU domain, the binding is not sequence specific and in such embodiments the vector does need not comprise a complementary binding motif.

Second Embodiment

A second illustrative embodiment of a plasmid according to the subject matter hereof is presented in FIG. 2 and a sequence for an embodiment of the vector is presented as SEQ ID NO: 2. This second embodiment is designated pFR1,5 SHT, reflecting the identity of the inserted hybrid protein.

The vector comprises gram-positive origin of replication, namely the PDOJHR origin of replication (2) whose sequence is comprised within SEQ ID NO: 28 and a gram negative origin of replication, namely the pUC ORI (1) presented as SEQ ID NO: 29. The vector comprises prokaryotic and eukaryotic expression cassettes. The prokaryotic expression cassette comprises a Ribosomal RNA promoter (8) and terminator (10) flanking coding sequences (9) for the hybrid protein SHT, defined elsewhere herein. The mammalian gene expression cassette

comprises a CMV promoter (4) and a TK Poly A site and terminator (6). In this case the inserted gene is a GFP reporter sequence (5).

In the example the plasmid comprises DNA sequences (11) suitable for binding to the sequence comprised in the SHT hybrid protein. The selectable marker in this plasmid is a spectinomycin resistance gene (3) but alternative selection markers, including those disclosed herein, will be readily identified and adopted by those skilled in the art. It will be understood that alternative forms of hybrid protein, alternative cargo sequences, and alternative protein binding motifs can readily be incorporated in variants of the illustrated embodiment. Likewise it will be understood that a wide range of well-known gram negative origins of replication will be selected amongst by those skilled in the art.

In a further aspect of the first embodiment there are disclosed hybrid proteins or fusion proteins. The hybrid proteins comprise at least one DNA binding domain, at least one cell penetrating peptide (CPP) domain, and at least one secretion signal sequence.

In embodiments there is disclosed a hybrid protein as described below or as otherwise disclosed herein. In embodiments the hybrid protein comprises: at least one signal sequence; at least one DNA binding domain; and at least one cell penetrating peptide (CPP) domain. It will be understood that in embodiments the protein comprises only one copy of a signal sequence domain, CPP domain and DNA binding domain but that in alternative embodiments any one, two or three of such domains may be present in more than one copy, or may comprise one, two, three or more different domains. Thus in embodiments, by way of example and not limitation, a protein may comprise multiple copies of the same DNA binding domain, or may contain copies of two, three or more different DNA binding domains. Similarly in embodiments the protein may comprise multiple copies of a CPP domain which copies may be the same or may be different. Similarly in embodiments the protein may comprise multiple signal sequences.

In embodiments the at least one CPP domain is or comprises a TAT domain, a VP22 domain, an Antp domain or a Rev domain. In alternative embodiments the CPP domain is or comprises a P-Beta MPG (gp41-SV40) domain, a Transportan (galanin mastoparan) domain, or a Pep-1 (Trp rich motif – SV40) domain. The sequences of such domains are presented herein as SEQ ID NOs 11 through 17.

In embodiments the at least one secretion signal sequence is selected from the group consisting of an alpha amylase signal sequence and an alpha arabinosidase signal sequence,

or a truncated form of such signal sequences. The sequences of these exemplary signal sequences and their corresponding DNA sequences are presented as SEQ ID NOs 18-23.

In embodiments the at least one DNA binding domain is or comprises at least one sequence specific DNA binding domain. In embodiments the at least one DNA binding domain comprises at least one domain selected from the group consisting of: a Zinc finger DNA binding domain, a homeobox DNA binding domain, a MerR DNA binding domain, and a HU DNA binding domain. DNA sequences encoding a selected DNA binding domain are presented as SEQ ID NOs 3 through 6.

Sequences of variant hybrid proteins comprising the foregoing domains are presented as SEQ ID NOs 24 through 27 and corresponding DNA sequences are presented as SEQ ID NOs 7 through 10. In embodiments the hybrid protein comprises an alpha arabinosidase signal sequence (SEQ ID NO: 23), a TAT domain (SEQ ID NO: 11) and a HU DNA binding domain (SEQ ID NO: 4).

In a further series of embodiments there is disclosed a method for transforming a target cell with a desired DNA, the method comprising the step of contacting the target cell with a protein-DNA complex comprising the desired DNA and the hybrid protein according to embodiments. In embodiments the method further comprises the step of contacting said DNA with said hybrid protein to form said protein-DNA complex. In embodiments the target cell is a eukaryotic cell and in embodiments is a prokaryotic cell and in embodiments is a gram positive prokaryotic cell.

In embodiments of the method the DNA is a plasmid and in embodiments the plasmid comprises an expression cassette for expressing the hybrid protein in a gram positive bacterium. In embodiments the plasmid comprises a eukaryotic expression cassette and in embodiments the eukaryotic expression cassette comprises inserted thereinto a DNA sequence for expression in a target eukaryotic cell. In embodiments the plasmid is a nucleic acid vector according to other embodiments disclosed herein.

In other embodiments of the method the at least one DNA binding domain is a sequence specific DNA binding domain and the plasmid comprises at least one nucleotide sequence for binding to said at least one sequence specific DNA binding domain. In embodiments of the method the at least one CPP sequence has at least 90% amino acid sequence identity over 10 contiguous amino acids to SEQ ID. NO. 11, 12, 13, 14, 15, 16, or 17 In embodiments of the method said at least one signal sequence has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%,

97%, 98%, 99% or 100% amino acid sequence identity over at least 10, 15, 20 or more contiguous amino acids to SEQ ID NO: 19, 21, or 23.

In embodiments of the method said at least one DNA binding domain has at least 0%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity over at least 10, 15, 20 or more contiguous amino acids to the amino acid sequence encoded by SEQ ID NOs 3, 4, 5 or 6.

In embodiments of the method the cell to be transformed is a gram positive bacterium and in embodiments is a eukaryotic cell which in embodiments is a mammalian cell and in embodiments is a human cell.

In related embodiments there is disclosed a cell transformed with an exogenous DNA using the hybrid protein according to embodiments wherein the cell is a gram positive bacterium or is a mammalian cell. In embodiments there is also disclosed the use of the hybrid protein according to embodiments hereof, to transform a gram positive bacterium or a mammalian cell. In embodiments the cell is a human cell. Examples of cells transformed using embodiments are disclosed elsewhere herein.

In embodiments the cell to be transformed using the hybrid protein is a *Staphylococcus, streptococcus, Bifidobacterium, lactococcus, lactobacillus, clostridium* or any other bacterial type disclosed herein. In embodiments the cell to be transformed is a mammalian cell and in embodiments is a bifidobacterium or a staphylococcus..

In embodiments there is disclosed a bacterial cell able to synthesize the hybrid protein according to embodiments. In embodiments the cell comprises a plasmid encoding the hybrid protein in a suitable expression context and in embodiments the cell is a bifidobacterium.

In embodiments the hybrid protein is synthsised by a bacterial cell and in alternative embodiments the hybrid protein is synthesized in silico using methods well known in the art and as further indicated herein.

KIT EMBODIMENTS

In a further series of embodiments there are disclosed kits for transforming a target cell with a desired DNA, the kit comprising a quantity of the hybrid protein according to embodiments, and instructions to: form a complex of the said hybrid protein with the desired DNA; and contact the said complex with said target cell.

In embodiments the kits comprise suitable media or buffers and solutions for said contacting or contain recipes or components for said media or buffers.

In embodiments the hybrid protein is encoded by a vector according to embodiments and is synthesized by a suitable host gram positive bacterium using a suitable expression cassette in the vector. In alternative embodiments the NHP is artificially synthesized using solid phase synthesis of peptide according to known techniques. In alternative embodiments the NHP is synthesized by a separate bacterium. In embodiments the purified or enriched NHP is added directly to a solution comprising a vector desired to be transformed into a target prokaryotic or eukaryotic cell, and the mixture containing the DNA/protein complex is contacted with the target cell. In embodiments such contacting of the protein and DNA occurs under suitable conditions for the formation of the complex to occur. While those skilled in the art will readily determine and optimize the conditions for the binding of particular combinations of hybrid protein and DNA using existing resources and the common general knowledge in the art, specific conditions used in non-limiting examples are presented in the Examples section hereof.

In embodiments of the NHP the DNA binding domain is one of a HU DNA binding domain, a Zinc finger DNA binding domain, a homeobox DNA binding domain, a homeobox DNA binding domain and a MerR DNA binding domain, or combinations thereof. Exemplary DNA sequences encoding suitable binding domains are presented as SEQ ID NOs 3, 4, 5, 6. In a range of alternative embodiments of the embodiments presented herein suitable DNA binding domains may be of any general type, including but not limited to helix-turn-helix, Zinc finger, leucine zipper, winged helix, winged helix turn helix, helix loop helix, HMG box, Wor 3 and RNA guided binding domains. Illustrative examples of DNA binding proteins whose DNA binding domains may be utilized in embodiments include histones, histone like proteins, transcription promoters, transcription repressors, transcriptional regulators, which may be drawn from a wide range of alternate sources and operons.

In embodiments the DNA binding domain is a HU DNA binding domain from the bacterial HU DNA binding protein. The sequence of the HU binding domain is presented at SEQ ID NO: 44 and its encoding DNA as SEQ ID NO: 4

In embodiments the CPP comprises a TAT domain (SEQ ID NO: 11), a VP22 protein of Herpes Simplex Virus (SEQ ID NO: 14), or the protein transduction domain of the Antennapedia (Antp) protein (SEQ, ID. NO. 12), or combinations thereof. Those skilled in the art will recognise that additional protein transduction domains can be used in alternative variants of the embodiment.

In one series of alternative embodiments the CPP domain comprises a Rev domain. Details of the foregoing and alternative possible transduction domains are described in Sugita et al., "Comparative study on transduction and toxicity of protein transduction domains" Br J Pharmacol. Mar 2008; 153(6): 1143–1152. FIG. 11 is a table showing the protein sequences of the four foregoing protein transduction domains and is taken from Sugita et al.

In embodiments the secretion signal sequence is alpha-L-arabinosidase (SEQ ID NO: 23), or a full length (SEQ ID NO: 19) or truncated (SEQ ID NO: 21)alpha-amylase signal sequence. The DNA sequences encoding the foregoing are presented as SEQ ID NO: 22, 18 and 20 respectively).

Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence (SEQ ID NO: 23), a Zinc finger DNA binding domain (SEQ ID NO: 46), and a TAT domain (SEQ ID NO: 11). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence (SEQ ID NO: 23), a Zinc finger DNA binding domain(SEQ ID NO:46), and a VP22 domain (SEQ ID NO: 14).Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence(SEQ ID NO: 23), a Zinc finger DNA binding domain(SEQ ID NO: 46), and the protein transduction domain of the Antp protein (SEQ ID NO:12). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence (SEQ ID NO: 23), a Zinc finger DNA binding domain. (SEQ ID NO: 46), Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence (SEQ ID NO: 23), a HU DNA binding domain, (SEQ ID NO: 44), and a TAT domain (SEQ ID NO: 11). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence (SEQ ID NO: 23), a HU DNA binding domain (SEQ ID NO: 44), and a V22 domain (SEQ ID NO: 14). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence (SEQ ID NO: 23), a HU DNA binding domain (SEQ ID NO: 44), and the protein transduction domain of the Antp protein (SEQ ID NO: 12).Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence(SEQ ID NO: 23), a HU DNA binding domain(SEQ ID NO: 44).

A) Variants based on alpha-L-arabinosidase

Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence (SEQ ID NO: 23), a Mer R DNA binding domain (SEQ ID NO: 41), and a TAT domain (SEQ ID NO: 11). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence a Mer R DNA binding domain (SEQ ID NO: 41), and a VP22 domain (SEQ ID NO: 14). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence a Mer R DNA

binding domain (SEQ ID NO: 41), and the protein transduction domain of the Antp protein (SEQ ID NO: 12). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence,a Mer R DNA binding domain(SEQ ID NO:41).

Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence (SEQ ID NO: 23), a HU DNA binding domain (SEQ ID NO: 44), and a TAT domain (SEQ ID NO:11). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence, a HU DNA binding domain (SEQ ID NO: 44), and a VP22 domain (SEQ ID NO: 14). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence, a HU DNA binding domain (SEQ ID NO: 44), and the protein transduction domain of the Antp protein (SEQ ID NO: 12). Thus, in embodiments, the NHP comprises an alpha-L-arabinosidase signal sequence, a HU DNA binding domain(SEQ ID NO: 44).

b) Variants based on alpha amylase signal sequence

Thus, in embodiments, the NHP comprises an alpha-amylase signal sequence (SEQ ID NOs 19, 21), a Zinc finger DNA binding domain (SEQ ID NO: 46), and a TAT domain (SEQ ID NO:11). Thus, in embodiments, the NHP comprises an alpha-L- amylase signal sequence (SEQ ID NOs 19, 21), a Zinc finger DNA binding domain (SEQ ID NO: 46), and a VP22 domain (SEQ ID NO: 14). Thus, in embodiments, the NHP comprises an alpha-L- amylase signal sequence (SEQ ID NOs 19, 21), a Zinc finger DNA binding domain (SEQ ID NO: 46), and the protein transduction domain of the Antp protein (SEQ ID NO: 12). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a Zinc finger DNA binding domain (SEQ ID NO: 46).

Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a homeobox DNA binding domain and a TAT domain (SEQ ID NO: 11). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs, 19, 21), a homeobox DNA binding domainand a VP22 domain (SEQ ID NO: 14). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a homeobox DNA binding domain, and the protein transduction domain of the Antp protein (SEQ ID NO: 12). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a homeobox DNA binding domain.

Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a Mer R DNA binding domain (SEQ ID NO: 45), and a TAT domain (SEQ ID NO: 11).

Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a Mer R DNA binding domain (SEQ ID NO: 45), and a VP22 domain(SEQ ID NO:14). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence, a Mer R DNA binding domain (SEQ ID NO: 45), and the protein transduction domain of the Antp protein (SEQ ID NO: 12). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence, a Mer R DNA binding domain (SEQ ID NO: 45). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a HU DNA binding domain (SEQ ID NO: 44), and a TAT domain (SEQ ID NO: 11). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence, a HU DNA binding domain (SEQ ID NO: 44), and a VP22 domain (SEQ ID NO: 14). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a HU DNA binding domain (SEQ ID NO: 44), and the protein transduction domain of the Antp protein (SEQ ID NO: 12). Thus, in embodiments, the NHP comprises an alpha-L-amylase signal sequence (SEQ ID NOs 19, 21), a HU DNA binding domain (SEQ ID NO: 44), and a CPP domain non limiting examples of which are presented as SEQ ID NOs 11 through 17.

In embodiments each component domains of the NHP is substantially full length. In embodiments each of the component domains of the NHP is a functional but partial domain.

Thus in embodiments there are disclosed the foregoing NHPs, as well as nucleic acid sequences encoding such NHP's and vectors comprising the nucleic acid sequences encoding the NHPs. Further in embodiments the encoding sequences are operatively linked to and transcribable by a host bacterium as part of a eukaryotic expression cassette.

In a series of embodiments of the first embodiment, the nucleic acid vectors comprise a cargo or candidate or exogenous gene or sequence is inserted into the eukaryotic expression cassette of the vector. The cargo gene may be any gene or nucleic acid sequence. In particular examples the gene is a marker gene such as GFP or is a tumor suppressor. It will be understood that the range of possible sequences that may be inserted into the eukaryotic expression cassette is not limited in any way. One skilled in the art will readily understand the adaptation and insertion of suitable sequences for expression in vectors according to this disclosure.

In embodiments the cargo or candidate sequence is a fluorescent or marker protein.

In embodiments the vector comprises at least one DNA motif for binding of a suitable DNA binding domain of a protein or of a hybrid protein according to an embodiment, has at least

90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO: 41 or 43 over at least about 10, 15, 20, or more contiguous nucleotides.

In embodiments the selection marker is effective in both gram positive and gram negative bacteria in embodiments the selection marker is resistance to spectinomycin. In embodiments the selection marker comprised in the vector is resistance to tetracycline. In embodiments the selection marker comprised in the vector is resistance to chloramphenicol. In embodiments the selection marker is any antibiotic effective against both gram negative and gram positive bacteria.

Third embodiment

In a third series of embodiments there are disclosed methods and compositions for transforming cells with DNA and expressing an exogenous or candidate or cargo DNA sequence in a target eukaryotic cell. In embodiments the target cell is a mammalian cell. In embodiments the target cell is a cancer cell. In embodiments the cancer cell is a lung or colon carcinoma cell and in embodiments is HEK-29 (human embryonic kidney); HT29, CACO (both human adenocarcinoma); HeLa, LL2 (lung carcinoma)Thus in one variant of the embodiment there is disclosed a method for transforming a eukaryotic target cell with a candidate DNA sequence, the method comprising the steps of: a) expressing in a first cell the hybrid protein from the first expression cassette of a nucleic acid vector according to the present invention, to form a complex between said hybrid protein and said nucleic acid vector; and b) contacting the target cell with the formed hybrid protein and nucleic acid vector complex, wherein said candidate DNA sequence is comprised in the second expression cassette.

In a further aspect of the embodiment there is disclosed a method for transforming a eukaryotic target cell with a candidate DNA sequence, the method comprising the step of: contacting the eukaryotic target cell with a nucleic acid vector complex formed by binding a vector according to the present invention with a hybrid protein comprising a signal sequence, a CPP sequence and a DNA binding sequence suitable to bind the vector.

In a further aspect of the embodiment there is disclosed a method for transforming a prokaryotic target cell with a candidate DNA sequence, the method comprising the step of: contacting the prokaryotic target cell with a nucleic acid vector complex formed by binding a vector according to the present invention with a hybrid protein comprising a signal sequence, a CPP sequence

and a DNA binding sequence suitable to bind the vector. In particular embodiments the hybrid protein-DNA complex is suitable to transform bifidobacteria, E. coli and staphylococcus.

Fourth series of embodiments

In alternative embodiments there are disclosed cells containing the nucleic acid vector according to any of the other embodiments. In embodiments the cell is a prokaryote a eukaryote, a mammalian cell, a human cell, a cancer cell, a probiotic bacterium, a gram negative bacterium or a gram positive bacterium.. In embodiments the cell is a bacterial cell, or is a gram positive or gram negative cell. In embodiments the cell is a human kidney cell, a human adenocarcinoma cell or a human lung carcinoma cell. In embodiments the cell is an HEK-29 cell (human embryonic kidney); an HT29, CaCo2 cell (both human adenocarcinoma); or an HeLa or LL2 (lung carcinoma) In embodiments a bacterial cell is *E. coli*, *bifidobacteria*, *lactococcus*, and *clostridium*, *staphylococcus* or *streptococcus*.

Fifth series of embodiments

In embodiments there are disclosed kits comprising vectors according to embodiments, or NHPs according to embodiments, or a combination of vectors and NHPs according to embodiments.

In embodiments there is disclosed a kit for transforming a cell with the vector according to claim 1, the kit comprising a quantity of the vector and a quantity of the vector and quantity of a hybrid protein comprising a DNA binding domain, a signal sequence domain and a CPP domain. In embodiments the kit further comprises a quantity of the nucleic acid vector.

In use the fusion protein according to an embodiment is used to transform a target prokaryotic cell or a target eukaryotic cell with a desired DNA. In embodiments the desired DNA is a plasmid or other vector. In embodiments the desired DNA comprises an expression cassette suitable to be expressed in the target prokaryotic cell.

In embodiments said at least one recognition nucleotide sequence has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97$, 98%, 99% or more sequence identity to SEQ ID. NO. 41 which is the MerR DNA motif or to SEQ ID NO: 43 which is the Zinc Finger DNA motif.

In embodiments the fusion protein comprises, sequentially: signal sequence-HU DNA binding domain-TAT domain; or signal sequence- Mer R DNA binding domain - TAT domain; or signal sequence- Zn finger DNA binding domain - TAT domain; or signal sequence- zinc finger DNA

binding domain - TAT domain. The nucleotide sequences encoding non-limiting examples of hybrid proteins according to embodiments are presented as SEQ ID NOs 7 through 10, and amino acid sequences are presented as SEQ ID NOs 24 through 27.

In embodiments said at least one recognition nucleotide sequence has at least about 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91% or 90% sequence identity to SEQ ID NO: 41 or 43 over a sequence of at least 10, 15, or 20 contiguous nucleotides.

EXAMPLE 1

PREPARATION OF NHP

NHP PREPARATION:

A NHP library was designed and created. This library was used to test hypothesis regarding the DNA binding, transfection and transformation capabilities of NHPs. The library consisted of: H (HU. SEQ ID NO: 44), M (Mer R SEQ ID NO: 45), L (Lac I SEQ ID NO: 47), Z (Zn finger SEQ ID NO: 46), and of combinations of a CPP domain and DNA binding domain HT (Hu-TAT combining SEQ ID NOs 44 and 11 in sequence), TH (Tat HU combining SEQ ID NOs 11 and 44 in sequence), LT (Lac Tat combining SEQ ID NOs 47 and 11 in sequence), TL (Tat – Lac I combining SEQ ID NOs 11 and 47 in sequence), TM (Tat – Mer R combining SEQ ID NOs 11 and 45 in sequence), MT (Mer R – Tat combining SEQ ID NOs 45 and 11 in sequence), TZ (Tat-Zn finger combining SEQ ID NOs 11 and 46 in sequence), ZT (Zn finger Tat combining SEQ ID NOs 46 and 1 in sequence, SHT (arabinosidase signa, Hu, Tat SEQ ID NO: 24), THS (Tat-Hu-arabinosidase signal) and SZT (arabinosidase signal sequence, Zn finger, Tat SEQ ID NO: 27). All proteins were chemically synthesized by solid phase peptide synthesis by LifeTein™ LCC.

The NHP library was used to characterize plasmid DNA binding abilities. It was demonstrated that all NHP's except those comprising a Lac-I DNA binding protein domain bound to plasmid DNA.

The different hybrid proteins display significant differences in binding affinities. Fig. 3 depicts the ability of SHT and SZT (SEQ ID NOs 24 and 27 respectively) to bind to a plasmid DNA (pBRA2.0 FIG. 1 and SEQ ID NO: 1). Fig. 3 shows that as concentrations of SHT (SEQ ID NO: 24) and SZT (SEQ ID NO: 27) increase, plasmid migration on a gel electrophoresis assay are retarded.

The NHP library was used to characterize plasmid DNA transfection into various mammalian cell lines including HEK-293, Hela, CaCo-2, LL2 and HT-29s. Various plasmids encoding the Green Fluorescent Protein (GFP) gene under control of the Cytomegalovirus promoter were bound to NHPs from the library. The NHP-bound plasmids were then incubated with the various cell lines, and examined under fluorescent microscopy to detect GFP expression. As expected, DNA binding domains alone did not result in plasmid transfection; however HT, LT, MT, ZT, SHT (SEQ ID NO: 24) and SZT (SEQ ID NO: 27) did result in plasmid transfection with ZT, SHT and SZT having the best transfection efficiencies. Reverse orientation NHPs such as TL, TZ, TS, TM and THS result in very limited plasmid transfection. This suggests that the TAT domain and DNA binding domains are necessary for plasmid transfection, specifically in the orientation with TAT domain on the amino-terminal. In addition, these findings also suggest that the secretion signal domain S, improves the transfection efficiency. Fig. 10 depicts SHT and SZT mediated transfection of the pBRA2.0 SHT plasmid (SEQ ID NO: 1) in HEK-293 and Hela cell lines. All other cell lines tested demonstrated consistent results.

A portion of the NHP library was used to examine the ability to transform bacteria with plasmid DNA. The NHP library included HT, ZT, TH, TZ, SHT and SZT. NHP-mediated transformation assays were tested on *Bifidobacteria longum, Staphylococcus aureus* and *Escherichia coli.* It was demonstrated that only SHT (SEQ ID NO: 24) and SZT (SEQ ID NO: 27) were able to transform *Bifidobacteria longum*, whereas HT, ZT, TH and TZ were unable to, suggesting that the secretion signal is necessary for this transformation ability. This experiment was then repeated with *Staphyloccocus aureus* and *Escherichia coli* confirming that SHT and SZT have cross-species transformation abilities. Compared to the traditional transformation method from gram-positive bacteria, electrotransformation, the use of SHT has been demonstrated to be superior, resulting in a higher rate of transformants. As SHT and SZT transform diverse bacterial species, it is believed that other bacterial species, both gram-positive and gram-negative will be able to effectively be transformed via this method. This includes species difficult to transform, similar to *Bifidobacteria longum,* which genera include, *lactococcus, clostridium, bacilius,* and *streptococcus.*

NHP-mediated gene delivery to both bacterial as well as mammalian target cells may therefore function through a cell-mediated internalization process. It is believed that the process requires the TAT domain and is enhanced with secretion signal domain. For bacterial cell transformation, it is believed that a species specific section signal plays a significant role in

transformation efficiencies, based on the reduced transformation efficiencies observed in *Staphylococcus aureus* and *Escherichia coli* assays compared to *Bifidobacteria longum.*

EXAMPLE 2

PLASMID PREPARATION

5 A DNA plasmid was designed and created to use *Bifidobacteria longum* as a vector for NHP-mediated gene delivery to mammalian cells (Figs. 1 and 2). The vector is comprised of 6 components that encode information allowing the *bifidobacteria* to carry out its designed function. The genetic components encoded include: (a) A mammalian expression cassette; (b) A prokaryotic expression cassette, encoding a novel hybrid protein; (c) A origin of replication in 10 *E. coli; (d)* An origin of replication in *bifidobacteria longum; (e)* A selectable marker; (f) Specific DNA binding sites for the novel hybrid protein. A more detailed explanation of the structure of the two vectors is presented above in the descriptions of embodiments.

Briefly, in the plasmid pBRA2.0 SHT (FIG. 1 and SEQ ID NO: 1) the first expression cassette for expression in a gram positive bacterium comprises an HU promoter and terminator flanking the 15 NHP encoding sequences as will be seen in FIG. 1. In pBRA2.0 SHT the second expression cassette for expression in a human cell, comprises the CMV (cytomegalovirus) promoter, a KOZAK sequence and a Thymidine Kinase polyadenylation/termination sequence, the foregoing flanking an insertion sequence for the insertion of a cargo sequence. This will be seen in SEQ ID NO: 1 wherein a GFP protein is inserted into the second expression cassette. In the 20 embodiment pBRA2.0 SHT the vector comprises a pUC origin of replication for replication in *E. coli* and a pB44 origin of replication for replication in *Bifidobacterium.* DNA-BS denotes a nucleotide binding site (SEQ.ID. NO. 43) for a Zinc finger DNA binding protein (SEQ ID NO: 46).

Plasmid pFRG1.5 SHT is shown in FIG. 2 and SEQ ID NO: 2 and the numbers refer to the numbering of features in FIG. 2. The vector comprises gram-positive origin of replication, 25 namely the PDOJHR origin of replication (2) whose sequence is comprised within SEQ ID NO: 28 and a gram negative origin of replication, namely the pUC ORI (1) presented as SEQ ID NO: 29. The prokaryotic expression cassette comprises a Ribosomal RNA promoter (8) and terminator (10) flanking coding sequences (9) for the hybrid protein SHT, defined elsewhere herein. The mammalian gene expression cassette comprises a CMV promoter (4) and a TK 30 Poly A site and terminator (6). In this case the inserted gene is a GFP reporter sequence (5). The plasmid comprises DNA sequences (11) suitable for binding by the sequence comprised in

the SHT hybrid protein. The selectable marker in the pFRG1.5 plasmid is a spectinomycin resistance gene (3).

Once the vector is inserted into the *bifidobacteria*, components (b), (d) and (f) are designed to activate due to interactions with the *bifidobacteria* cell machinery. As a result, the novel hybrid protein is expressed, binding to a specific binding site on the vector, and results in the entire vector-novel hybrid protein complex being secreted out of the *bifidobacteria* and delivered into mammalian cells. Once in the mammalian cell, the mammalian expression cassette component activates due to interactions with the cell machinery and it expresses the therapeutic gene to carry out a desired function. Components (c) and (e) are not therapeutically relevant but are required during the construction and validation of the technology.

The function of each component was tested experimentally after having the vector synthesized from GeneArt Inc.

A mammalian gene expression cassette

The ability of the vector to express the mammalian gene expression cassette was examined in various mammalian cell culture lines including LL2, Hela, Hek-293, HT-29 and CaCo-2 cells. A mammalian expression cassette that contained the Green Fluorescent Protein gene was used, allowing the use of fluorescent microscopy to detect proper function of the genetic component. Fig. 5 depicts pBRA2.0 SHT (SEQ ID NO: 1) transfection in HeLa and HEK-293 cell lines, and resulting mammalian expression cassette function.

A prokaryotic gene expression cassette

The ability of the vector to express the prokaryotic mammalian gene expression cassette in *Bifidobacteria longum* was examined. To do this, immunoblot analysis on cell lysate of *Bifidobacteria longum was performed.* Fig. 9 depicts the immunoblot, with a band being detected at the predicted NHP size of 14 kDa in lane 9.

An origin of replication in E.coli and selectable marker

To test this component, E.coli cells were transformed with pBRA2.0 SHT (SEQ ID NO: 1). After propagating them, plasmid were purified followed by gel electrophoresis to confirm positive transformants and the function of the origin of replication. Fig. 4 depicts the gel electrophoresis of pBRA2.0 SHT (SEQ ID NO: 1, FIG. 1) purified from E.coli and confirms positive transformants. After transformation the culture was propagated in spectinomycin containing

culture broth. Propagation in this broth suggests the plasmid is being replicated and the selectable marker functions.

An origin of replication in bifidobacteria longum and selectable marker

To test this component, bifidobacteria cells were transformed with pBRA2.0 SHT (SEQ ID NO: 1, FIG. 1) purified from E.coli. After transformation the culture were propagated in spectinomycin containing culture broth. Propagation in this broth suggests the plasmid is being replicated and the selectable marker functions.

A specific DNA binding site for NHPs

To test this function, purified pBRA2.0 SHT (SEQ ID NO: 1, FIG. 1) was used and bound with SHT and SZT (SEQ ID NOs 24 and 27) proteins. Fig. 3 depicts this binding assay with pBRA2.0 SHT migration being retarded with increasing NHP concentrations. Additionally, SHT and SZT mediated pBRA2.0 SHT delivery to various cell lines were performed. The positive result is the expression of the mammalian gene expression cassette encoding GFP. Fig. 5 depicts GFP positive cells post NHP-mediated pBRA2.0 SHT (SEQ ID NO: 1, FIG. 1) transfection.

Collective Vector function:

After validating the components of the vector independently their collective function as a gene delivery system were examined. To do so, NHP-bound pBRA2.0 SHT (SEQ ID NO: 1, FIG.1) was screened for in the supernatant of *Bifidobacteria longum* cultures.

The first step was to detect pBRA2.0 SHT (SEQ ID NO: 1, FIG. 1) plasmid DNA via PCR (Fig. 6). After rigorous centrifugation, supernatants of wild-type bifidobacteria, pTM13 positive transformants and pBRA2.0 SHT positive transformants were collected. pTM13 is the same vector as pBRA2.0 SHT without the prokaryotic gene expression cassette encoding NHP, as a result we anticipated no plasmid secretion in pTM13 transformants. After the PCR, positive controls confirmed the validity of the PCR reaction, whereas negative control, including wild-type and pTM13 positive transformants did not have template amplification. Only pBRA2.0 SHT positive transformants provided amplified template from isolated supernatant.

Second, pBRA2.0 SHT (SEQ ID NO: 1, FIG. 1) was isolated from the supernatant, using plasmid purification protocols on the collected supernatant that included modified Qiagen™ prep kits, as well as traditional phenol-chloroform purification protocols (Figs. 7 and 8). The figures

depict an isolated vector corresponding to the size of pBRA2.0 SHT (SEQ ID NO: 1, FIG. 1), and upon restriction digest analysis, was confirmed to be pBRA2.0 SHT.

It is believed that increasing the expression of NHP and the copy number of the vector will provide sufficient NHP-bound vector to detect NHP-mediated gene delivery in mammalian cells. To do this, a vector having enhanced versions of component (b) and component (d) such as pFRG1.5 (SEQ ID NO: 2, Fig. 2). It is also believed that changing the components to become species specific, for another species is relatively straight forward, as other gene expression cassettes, origins of replications, and selectable markers can be identified through genomic analysis. Based on the current designs, it is believed that the vector can be specifically designed and tested to work in species similar to *Bifidobacteria* such as *lactococcus, clostridium, staphylococcus, streptococcus* and *bacilius.*

EXAMPLE 3

CULTURE TECHNIQUES, SERIAL PASSAGES, PLATING AND STORAGE OF BIFIDOBACTERIAL CELLS

Bifidobacterium longum cells were subcultured in screw capped anaerobic tubes containing 7-7.5 mL RCB (Reinforced Clostridial broth). After 18-24 hours of incubation, the cells were plated on RCBA (Reinforced Clostridial agar) and incubated for 24-48 hours at 37°C under anaerobic conditions. A single colony was picked and again subcultured in screw capped anaerobic tubes containing 7-7.5 mL RCB. This procedure of 2 serial subcultures or passages (10% V/V) was used as a starter culture for all the experiments. After the cells reach an $OD_{600\ nm}$ of 1.4-1.8, cell pellets were collected and frozen at -80°C for future use. Also, a glycerol stock of 2 mL of active culture is added to 0.5-1 mL 50% of sterile glycerol in a cryo vial and stored at -80°C. To obtain an active bacterial cell culture, the cells were constantly passaged every 2 days in sterile RCB tubes.

EXAMPLE 4

ELECTROTRANSFORMATION IN BIFIDOBACTERIUM LONGUM

Preparation of cells for electroporation

Bifidobacterium longum overnight culture (10% V/V) was inoculated in fresh Reinforced Clostridial broth (RCB) in 10 mL screw capped anaerobic tubes and incubated at 37°C under anaerobic conditions. This culture was serially passaged in sterile RCB tube again under

anaerobic conditions at 37°C until an $OD_{600\ nm}$ of 0.6-0.7 was reached. Once the cells reached mid-exponential phase, the cells were first chilled on ice and centrifuged at 4,500 rpm for 15 min at 4°C in a 15 mL Falcon tube. After centrifugation, the cell pellet was resuspended and washed twice with ice-cold 0.5 M sucrose in eppendorf tubes. After, the cells were resuspended in 1/250 of the original culture volume (10 mL), which is 300 µL of ice-cold electroporation citrate buffer consisting of 1 mM ammonium citrate + 0.5 M sucrose (pH- 6.) at 4°C for 2.5 hours.

Electroporation protocol

Plasmid DNA (400, 600, 800, 1000 ng: 800 ng best) were mixed with 80 µL cell suspension in a precooled Bio-Rad™ Gene Pulser™ disposable cuvette with an interelectrode distance of 0.2 cm. A Bio-Rad™ Gene Pulser™ was used to deliver a high-voltage electric pulse of 2000 V along with 25 µF capacitance and 200 Ω resistance. Following electroporation, the transformant mixture was transferred in eppendorf tubes containing 900-920 µL of sterile RCB containing 0.5 M sucrose and the cells were incubated at 37°C for 3-4 h inside the anaerobic glove bag. This procedure is required for cell recovery and the expression of antibiotic resistance marker. After cell recovery, the cells were subcultured on sterile RCB containing 25, 50, and 100 µg/mL of Ampicillin and Spectinomycin antibiotics respectively and incubated at 37°C for 2-4 days.

Selection of transformants

For selection, RCBS and RCBAS (Reinforced Clostridial medium with spectinomycin for vector PTM13) were used for selection of transformants. The concentrations used for selection of positive transformants were 50 µg/mL, 75 µg/mL, 100 µg/mL, 125 µg/mL and 150 µg/mL spectinomycin in tubes and agar plates. For preparing liquid selective media, the above-mentioned concentrations of spectinomycin were added to sterile RCBS tubes inside the anaerobic glove bag asceptically. For selection of transformants after electroporation, 50-100 µL of transformant mixture was plated using spread plate technique on to sterile RCAS agar plates pre-spread with appropriate concentrations of spectinomycin and incubated for 48-72 h at 37°C inside the glove bag. Control plates containing the same concentrations of spectinomycin were used for both sterile tubes and agar plates containing (10% V/V for liquid broth and 50-100 µL of culture for agar plates) wildtype *Bifidobacerium longum* cells. After 48-72 hours or even further sometimes, the positive transformant colonies were picked using a sterile loop inside the glove bag and subcultured in tubes containing selective RCBS with 125 µg/mL spectinomycin. No

growth was observed on the control RCBAS plates with the same concentrations of spectinomycin even after a week.

Screening positive transformants:

Plasmid Isolation

5 Harvest the cells (6-7 mL culture) at 5,000 rpm at 4°C for 15 min after overnight growth from single colonies from the selective agar plates in 15 mL Falcon tubes. Wash the pellets with 1 mL of PBS once to remove excess end products in the medium. Resuspend pellets with 570 µL of solution A {6.7% sucrose, 50 mM Tris-HCl, 1 mM EDTA (pH-8.0)} and prewarm at 37°C for 5 min. Cell lysis was performed by the addition of 145 µL of solution B {25 mM Tris-HCl (pH-8.0),
10 20 mg/mL Lysozyme} and incubated at 37°C for 40-45 min. Further, 72 µL of solution C {0.25 mM EDTA, 50 mM Tris-HCl (pH-8.0)} and 42 µL of solution D {20% SDS, 50 mM Tris-HCl, 20 mM EDTA (pH-8.0)} were added, mixed gently and incubated for 10 min at 37°C. After incubation, the tubes were mixed briefly by inverting for few seconds and the genomic DNA was denatured by adding 42 µL of 3M NaOH and gently mixing for 10 min. The solution was
15 neutralized by adding 75 µL of 2 M Tris-HCl (pH-7.0) for 3 min and the DNA was precipitated using 107 µL of 5 M NaCl for 1 min by gentle mixing. The precipitated genomic DNA was removed by centrifugation at 14,000 rpm for 10-15 min at 4°C. The supernatant was added to a new eppendorf tube containing 700 µL of phenol:chloroform:isoamyl alcohol mixture (25:24:1) and mixed vigorously. Following that, the mixture was centrifuged at 14,000 rpm for 15-20 min
20 at room temperature. After centrifugation, the upper aqueous phase was transferred to a new tube by careful aspiration or gentle pipetting and treated with 600 µL of isopropanol and the DNA was precipitated for 30 min on ice. Then, the tubes were centrifuged at 14,000 rpm for 15-20 min at 4°C. After centrifugation, the supernatant was discarded and the pellets were washed with ice-cold (-20°C) 70% ethanol.

25 After centrifugation at 14,000 rpm for 10 min, the supernantant was discarded and the pellets were air-dried for 5-10 min. The plasmid DNA pellets were resuspended in 35 µL of nuclease free water or Tris-HCl buffer. 1.5 µL of RNAse was added and incubated at 37°C for 15-30 min.

Restriction Digestion and agarose gel electrophoresis

17 µL of Plasmid DNA (550-600 ng) were added to a mixture containing 2 µL of EcoR1, 4 µL of
30 reaction buffer, and 17 µL of water. The digestion mixture was incubated at 37°C for 2-2.5 h.

Then, the reaction was heat inactivated at 65°C for 10-15 min. The samples including transformant (cut, uncut) plasmid DNA, control: wildtype grown in the absence of antibiotic (cut, uncut), control: wildtype grown in the presence of antibiotic (cut, uncut), vector DNA (PTM13) cut/uncut, and DNA ladders were loaded on 1.5% agarose gel and electrophoresed to see the appropriate band patterns.

Other Plasmid Isolation Protocols

Buffers needed: TES (25 mM Tris-Cl, 10 mM EDTA, 50 mM sucrose); KOAc: 3 M K^+, 5 M acetate (3 M potassium-acetate, 2 M acetic acid -- glacial is 17 M); PBS, phosphate buffered saline pH 7.0.

1mL of fresh culture grown anaerobically at 37°C with mid-log phase OD (~8-10 h) was used for plasmid isolation. Centrifuge 20 min at 4,000 RPM at 4°C in centrifuge. Discard the supernatant, and wash in 5 mL of PBS buffer (pH 7.0). Centrifuge again and discard supernatant. Add 300-400 µL TES containing 6 mg/mL lysozyme and 40 µg/mL mutanolysin and resuspend the pellet in a 2 mL eppendorf tube. Incubate at 37°C for 60 minutes. With O'Sullivan's method, used 200 µL of 25% sucrose with 30 mg/mL lysozyme as resuspension buffer and incubated for 15-20 min at 37°C. Meanwhile, 0.2 N NaOH and 2% SDS was freshly made to perform lysis. Add 600 µL SDS/NaOH mix to each tube. Incubate on ice for 5 min. With O'Sullivan's method, also tried with 400 µL of lysis buffer (3% SDS and 0.2 N NaOH) and mixed by inversion and incubating at room temperature for 5-7 min. Add 500 µL KOAc (3M K^+, 5M acetate). Incubate on ice for 2 min. Shake vigorously and centrifuge at 17,200 g/12,000 rpm, 4°C, for 15 min. With O'Sullivan's method, used 300 µL of ice-cold 3 M sodium acetate (pH-4.8) and mixed by inverting the tube and centrifuged at 13,500 rpm for 15 min at 4°C. Remove supernatant (~1 mL) into a fresh 2 mL tube and discard the pellet and remaining supernatant. Add ½ volume (500 µL) of isopropanol. Centrifuge at 17,200 g/12,000 rpm (room temperature) for 10 min. Discard the supernatant. With O'Sullivan's method, the supernatant was collected in a new eppendorf and mixed with room temperature 650 µL isopropanol and centrifuged at 13, 500 rpm for 10 min at 4°C. Wash pellet with 1mL 70% ethanol. Air dry for 2-5 min on bench. Resuspend in 500 µL TE buffer. Add 500 µL 5M LiCl. Incubate on ice for 5 min. Centrifuge at 17,200 g/12,000 rpm for 10 min. Pour supernatant into an Eppendorf. Add 1mL isopropanol. Incubate on the bench for 10 min. Centrifuge at 17,200 g/12,000 rpm for 10 min. Discard the supernatant. Wash the pellets with 100 µL 70% ethanol. Resuspend in 375 µL TE buffer. Add 5 µL 1 mg/mL RNase A. Incubate at 37°C for 30 min. The same protocol was performed without LiCl step after 70% ethanol wash,

the air-dried pellets were resuspended in TE and subjected to RNAase A treatment. With O'Sullivan's method, the DNA pellets were resuspended in 500 µL of sterile nuclease free water after isopropanol step and then subjected to phenol-chloroform purification. Add 700 µL phenol:chloroform:isoamyl alcohol. Vortex until thoroughly mixed. Centrifuge at top speed of microfuge for 2 min. Pipette aqueous phase (the top one) into new Eppendorf. Then repeat the procedure twice with the same volume of chloroform:isoamyl alcohol to remove any phenol. Add 750 µL straight ethanol and 125 µL 3M sodium acetate. Put at -80°C for 30 min or -20°C overnight. With O'Sullivan's method, the upper phase was mixed with 1 mL of ethanol (-20°C) and centrifuged at 13,500 rpm for 15 min at 4°C. Centrifuge at 13,600 g/12,000 rpm, 4°C, for >15 min. Discard the supernatant. Wash pellet with ~100 µL 70% ethanol. Resuspend in 200 µL Tris buffer per 40 mL of original cells. With O'Sullivan's method, the DNA pellets were washed with 70% ethanol, air-dried and resuspended in 50 µL of TE containing RNAse A.

EXAMPLE 5

PBRA2.0 SHT

The structure of pBRA2.0 SHT is shown in FIG. 1 and as SEQ ID NO: 1. pBRA2.0 SHT and its nucleotide sequence is presented as SEQ ID NO: 1 The vector is a plasmid and as indicated in FIG. 1 it comprises a spectinomycin resistance gene SmR, an *E. coli* pUC origin of replication and a *Bifidobacterium longum* origin of replication derived from pB44. The eukaryotic expression cassette comprises a CMV (cytomegalovirus) promoter, and a Thymidine kinase terminator and polyadenylation site from the HSV virus. A candidate or cargo sequence (designated as "GOI") can be inserted in the correct reading frame for expression through operation of the CMV promoter.

The vector also comprises an expression cassette for expression of a cargo fusion protein or NHP. As will be seen in the Figure, the prokaryotic expression cassette is framed by the Hu promoter HuP and terminator HuT. The desired NHP open reading frame is inserted between the two in the appropriate reading frame and under the control of the Hu promoter.

The HuT sequence is followed by the DNA-BS sequence which is a DNA motif bound to by a Zinc finger DNA binding domain. It will be understood that in different variants the protein chosen may have the sequence SHT, SZT, SMT, SLT (SEQ ID NOs 24, 25, 26, 27). In the examples presented here the NHP has the SHT structure, the coding nucleotide sequence for which is shown as SEQ ID NO: 24, and its corresponding DNA sequence as SEQ ID NO: 8.

In this embodiment the Signal sequence is alpha arabinosidase, SEQ ID NOs 22, 23, the DNA binding domain is the Hu domain the DNA sequence for which is presented as SEQ ID NO: 4 and the CPP domain is the TAT domain SEQ ID NO: 11. The pBRA 2.0 plasmid sequence was assembled from individual DNA sequences and synthesized commercially by LifeTein™ LCC.

5

EXAMPLE 6

pFRG1.5

The structure of pFRG1.5 SHT is shown in FIG. 2 and its nucleotide sequence is shown as SEQ ID NO: 2

The vector comprises gram-positive origin of replication, namely the pDOJHR origin of
10 replication whose sequence is presented as a part of SEQ ID NO: 28 and a gram negative origin of replication, namely the pUC ORI (E. coli ORI) presented as SEQ ID NO: 29. The vector comprises prokaryotic and eukaryotic expression cassettes. The prokaryotic expression cassette comprises a ribosomal RNA promoter and terminator flanking coding sequences fort the hybrid protein SZT, defined elsewhere herein. The mammalian gene expression cassette
15 comprises a CMV promoter and a TK Poly A site and terminator. In this case the inserted gene is a GFP sequence.

In the example the plasmid comprises DNA sequences suitable for binding a Zinc finger consensus sequence. The selectable marker in this plasmid is a spectinomycin resistance gene. pFRG was were obtained commercially and was synthesized using standard methods.

20

EXAMPLE 7

VALIDATION OF VECTOR

In the example a pBRA2.0 SHT plasmid containing the GFP marker was transformed into bifidobacterial cells. The bifidobacterial cells were confirmed as hosting the plasmid, and expressing the encoded NHP protein. Secretion of the pBRA2.0 plasmid from the bifidobacterial
25 cells was confirmed directly.

pBRA2.0 SHT *E. coli* transformation assay to validate the function of the *E. coli* ORI

pBRA2.0 SHT was purified from transformed *E. coli* DH5α chemically competent cells to validate the *E. coli* ORI on pBRA2.0 SHT. Positive transformants were selected for on agar plates containing 100 μg/μL of spectinomycin. Cells were subject to plasmid purification using

Qiagen QIAprep Spin Miniprep Kit following manufacturers protocols and subject to restriction enzyme digest using EcoR1 and visualized on a 0.8% agarose gel.

FIG. 4 shows the results of running samples of the resulting DNA digests on an agarose gel. Lanes are as follows: **1.** 1kb ladder, **2.** Chemically synthesized vector, **3.** pBRA2.0 SHT positive transformant A, **4.** pBRA2.0 SHT positive transformant B.

pBRA2.0 SHT transfection assay, validation of the eukaryotic ORF

pBRA2.0 SHT (SEQ ID NO: 1, FIG.1) was subject to transfection for validation of eukaryotic open reading frame on pBRA2.0 SHT. pBRA2.0 SHT was transfected into human cell lines using Lipofectamine™ (Life Technologies™) following manufacturers protocols. Cells were visualized using direct fluorescence microscopy for GFP expression. The results are shown in FIG. 5: Panel 1 HEK-293 cells: A. control; B. pBRA2.0 SHT; C. pBRA2.0 SHT + Lipofectamine™. Panel 2 HeLa cells: A. control; B. pBRA2.0 SHT; **C.** pBRA2.0 SHT + Lipofectamine™.

A comparison of the experimental panels 1C and 2C of Fig. 5 with the controls demonstrates the expression of GFP from the eukaryotic expression cassette in these cells.

Validation of secretion of vector. PCR screening of pTM13 and pRBA2.0 supernatants, validation of pBRA2.0 SHT secretion

Cultures were grown to OD 1.8 where 1 mL of supernatant for wild-type, pTM13 (a plasmid not containing the SHT sequence) and pBRA2.0 SHT cultures were subjected to plasmid purification using Qiagen™ QIAprep™ Spin Miniprep Kit following manufacturers protocols. Resulting eluates were subject to PCR using pTM13/pBRA2.0 SHT specific primers to validate plasmid secretion from pBRA2.0 SHT transformants. A series of primer pairs useable to identify the presence of the plasmid are presented as SEQ ID NOs 25 and 36 (primer pair to amplify GFP sequences); SEQ ID NOs 37 and 38 (primer pair to amplify spectinomycin resistance gene sequences); and SEQ ID NOs 39 and 40 (primer pair to amplify CMV sequences). As a further confirmation of the results, the amplicons were sequenced to confirm sequence identity by at the Center for Molecular Medicine and Therapeutics at the DNA Sequencing Core Facility in Vancouver B.C. Results are shown in FIG. 6. 1: 1 kb Ladder; 2: 100 bp Ladder; 3: PCR negative control-GFP primers, no template; 4: PCR negative control-SpecR primers, no template; 5: PCR negative control-CMV primers, no template; 6: PCR positive control-GFP, pBRA-SHT plasmid. DNA; 7: PCR positive control-SpecR, pBRA-SHT plasmid. DNA; 8: PCR

positive control- CMV, pBRA-SHT plasmid DNA; 9: PCR positive control-GFP, pTM13C plasmid DNA; 10: PCR positive control-SpecR, pTM13C plasmid DNA; 11: PCR positive control-CMV, pTM13C plasmid DNA; 12: Control *B. longum* P. DNA-GFP; 13: Control *B. longum* P. DNA-SpecR; 14: Control *B. longum* P. DNA-CMV; 15: PTM13-transformant supernatant-P. DNA-GFP; 16: PTM13-transformant supernatant- P. DNA-SpecR; 17: PTM13-transformant supernatant- P. DNA-CMV; 18: pBRA-SHT transformant supernatant- P. DNA-GFP; 19: pBRA-SHT transformant supernatant-P. DNA-SpecR; 20: pBRA-SHT transformant supernatant- P. DNA-CMV.

Thus the results of this experiment demonstrate that pBRA vector is secreted by the bacteria, whereas the pTM13 plasmid is not.

Visualization and digestion of pBRA2.0 SHT supernatant, validation of pBRA2.0 SHT secretion and characterization

pBRA2.0 SHT culture was grown to OD1.8 where 1 mL of supernatant of culture was subjected to plasmid purification using Qiagen™ QIAprep™ Spin Miniprep Kit following manufacturers protocols. Eluate was subject to restriction enzyme digest and visualized on a 0.8% PAGE to validate pBRA2.0 SHT secretion from positive transformants. The results are shown in FIG.7. Panel A: 1. 1kb ladder; 2. pBRA2.0 SHT; 3. pBRA2.0 SHT from pDNA Miniprep Kit. Panel B: 1. 1kb ladder; 2. pBRA2.0 SHT digested with EcoR1; 3. pBRA2.0 SHT from pDNA Miniprep Kit digested with EcoR1.

The results of this experiment more fully confirm the secretion of pBRA2.0 SHT from the bacterial cells.

Visualizations and digestions of concentrated supernatants, validation of pBRA2.0 SHT secretion and characterization

Wild-type, pTM13, pBRA2.0-SHT (pBRA comprising SHT SEQ ID NO: 24 in the prokaryotic expression cassette and pBRA2.0-SZT (comprising instead the SZT protein SEQ ID NO: 27) cultures were grown to an OD of 1.8. 7 mL of supernatant were concentrated under each of the aforesaid conditions using 3000 MWCO concentrators and samples were subject to phenol-chloroform plasmid DNA extraction. Results were subject to restriction enzyme digest using EcoR1 and visualized on a 0.8% PAGE to valid DNA secretion in pBRA2.0 SHT transformants. Results are shown in FIG. 8. Panel A – Phenol-chloroform samples: 1. 1kb ladder; 2. 100 bp ladder; 3. Blank; 4. Wild-type, 5. pTM13 transformants, 6. Blank; 7. pBRA2.0-SHT; 8. pBRA2.0-

SZT. Panel B: -Phenol-chloroform samples digested with EcoR1. 1. 1kb ladder; 2. Wild-type; 3. Blank; 4. pTM13 transformants; 5. Blank; 6. pBRA2.0-SHT transformants; 7. Blank; 8 pBRA2.0-SZT transformants.

EXAMPLE 8

WESTERN BLOT ANALYSIS OF SZT, SHT VALIDATION OF *BIFIDOBACTERIUM* SPECIFIC ORF

Wild-type, and pTM13, pBRA2.0-SHT and pBRA2.0-SZT transformed cultures were grown to an OD of 1.8. Samples were loaded onto an 0.8% PACM gel and subject to electrophoresis, and transferred to a membrane for immunoblot analysis using anti-TAT antibodies to detect the expression of our hybrid proteins SHT and SZT SEQ ID NO: 27, to validate the functionality of the *Bifidobacterium* specific promoter and terminator. Samples were prepped from freshly grown cells washed three timex in sterile PBS. Cells were pelleted and subject to sonication in deionoized water with three pulses of 20 seconds with 30 second intervals in an ice bath. Samples were quantified using via Bradford and aliquots of each sample were mixed with 4x SDS loading dye and boiled at 95 degree Celsius for 10 minutes and quenched on ice. These results suggest the promoter and terminator functionality of the HU design on pBRA2.0 is sufficient for protein expression. Results are shown in FIG. 9. **1.** Precision Plus Protein™ Prestained Standards (Bio-Rad); 2. Wild-type, 3. Wild-type; 4. pTM13; 5. pTM13; 6. pBRA2.0-SHT; 7. pBRA2.0-SHT; 8. pBRA2.0-SZT; 9. pBRA2.0-SZT.

SHT and SZT interact with pBRA2.0 SHT and result in gel-shift with increasing concentrations of peptide

pBRA2.0 SHT was incubated with SHT or SZT SEQ ID NO: 27 in binding buffer at room temperature for 15 minutes and subject to PAGE. Results are shown in FIG. 3. 1. 1 kb ladder; 2. 100 bp ladder; 3. 500 ng of pBRA2.0 SHT; 4. 500 ng of pBRA2.0 SHT + 2 ng of SHT; 5. 500 ng of pBRA2.0 SHT + 5 ng of SHT; 6. 500 ng of pBRA2.0 SHT + 10 ng of SHT; 7. 500 ng of pBRA2.0 SHT + 20 ng of SHT; 8. 500 ng of pBRA2.0 SHT + 30 ng of SHT, 9. 500 ng of pBRA2.0 SHT + 40 ng of SHT; 10. 500 ng of pBRA2.0 SHT + 50 ng of SHT; 11. Blank; 12. 500 ng of pBRA2.0 SHT; 13. 500 ng of pBRA2.0 SHT + 2 ng of SZT; 14. 500 ng of pBRA2.0 SHT + 5 ng of SZT; 15. 500 ng of pBRA2.0 SHT + 10 ng of SZT; 16. 500 ng of pBRA2.0 SHT + 20 ng of SZT; 17. 500 ng of pBRA2.0 SHT + 30 ng of SZT, 18. 500 ng of pBRA2.0 SHT + 40 ng of SZT; 19. 500 ng of pBRA2.0 SHT + 50 ng of SZT; 20. Blank.

EXAMPLE 10

COMPLEXES OF PBRA2.0 SHT DNA AND SHT PROTEIN AND OF PBRA2.0 SHT DNA AND SZT PROTEIN CAN TRANSFECT AND EXPRESS GFP IN MAMMALIAN CELL LINES

SHT and SZT SEQ ID NO: 27 were subject to DNA binding assay followed by transfection assay to validate the cell-penetrating functional domain of each peptide. pBRA2.0 SHT was transfected into human cell lines using Lipofectamine (Life Technologies) following manufacturers protocols as a positice control. Cells were visualized using direct fluorescence microscopy for GFP expression. The results are shown in FIG. 10. Panel 1 – HEK-293 cells: A. control; B. pBRA2.0 SHT; C. pBRA2.0 SHT + Lipofectamine; D. pBRA2.0-SHT complex Pane 2 – HeLa cells: A. control; B. pBRA2.0 SHT; C. pBRA2.0 SHT + Lipofectamine; D. pBRA2.0-SZT complex.

EXAMPLE 11

TRANSFORMATION OF CELLS USING HYBRID PROTEIN

A hybrid protein comprising arabinosidase signal, HU DNA binding domain and Tat transduction sequence, was bound to pBRA2.0 SHT plasmid DNA in binding buffer (150mM NaCl 50mM Tris pH 7.2) Unless otherwise set out below, procedures were the same as set out for Example 1 above. Protocol for the Hybrid Protein-pBRA2.0 Binding Assay were as follows:

1. Aliquot 1 ng of SHT and 150 ng of pBRA2.0 1 into a 1.5 mL Eppendorf tube containing 50 µL of NHP Binding Buffer (50mM NaCl 50mM Tris pH 7.2).

2. Repeat step one, while increase concentration of SHT independently by 10, 20, 30, 50, 100, 500 and 1000 ng for resultant laddering effect to determine concentration dependent Novel Hybrid Protein-pBRA2.0 Binding.

3. Leave at room temperature for 30 minutes and visualize on a 0.8% agarose gel electrophoresis.

Protocol for the Hybrid Protein-pBRA2.0 Transformation Assay:

1. Aliquot 5 µg of SHT and 400 ng of pBRA2.0 1 into a 1.5 mL Eppendorf tube containing 50 µL of NHP Binding Buffer (50mM NaCl 50mM Tris pH 7.2 and leave at room temperature for 30 minutes.

2. Resuspend 150 µL of $5x10^8$ cells/mL into reaction mixture (Step 1).

3. Under anaerobic conditions, gently rock tube for 30 minutes containing protein-DNA complexes.

4. Add 800 µL of RCB for cell recovery under anaerobic conditions, incubate at 37°C for 4 hours.

5. Plate 80 µL of reaction mixture on RCA-spectinomycin (200 µg/mL) resistance plates and incubate for 3 days at 37°C under anaerobic conditions.

6. Screen for positive transformants.

SEQUENCE LISTINGS

The following sequence listings are incorporated herein and form an integral part of this disclosure.

SEQ ID NO: 1: pBRA2.0-SHT

CCAGGCCCGTGGAGGCGAGGAAGACGGACGGCGACGGCAAGGGCCATTGGACGAGCGT
GGCGGGGTATGGCGAGGTGTTCACGACCACGGAGCTGTTCGACGTGACGGCCGCGCGTG
ACCACTTCGACGGCACCGTGGAGGCCGGGGAATGCCGTTTCTGCGCGTTTGACGCGCGC
AACCGCGAACATCATGCGCGGAACGCCGGAAGGTTGTTCTAGCGGCCGTGTCCGCGCCT
CTGGGGCGGTTGCGCCTGCCATGGAGATCTGGGGCCGAGTCGGCCGCGGGCTTCGAGG
GAGGCGACGAGAGCACATCGCCCGCCTCAGGCGACGAGAGCACATCGCCCGCCTCGGTC
GGCCGCGAGGCGACGAGAGCACATCGCCCGGGCCGAGTCGGCCGCGGGCTTCGAGGGA
GGTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTT
CATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC
ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAAG
AAGGATGCTTTCTCGAGATGACCCTGACCGGCACCCTGCGCAAAGCGTTTGCGACCACCC
TGGCGGCGGCGATGCTGATTGGCACCCTGGCGGGCTGCAGCAGCGCGGCATACAACAAG
TCTGACCTCGTTTCGAAGATCGCCCAGAAGTCCAACCTGACCAAGGCTCAGGCCGAGGCT
GCTGTTAACGCCTTCCAGGATGTGTTCGTCGAGGCTATGAAGTCCGGCGAAGGCCTGAAG
CTCACCGGCCTGTTCTCCGCTGAGCGCGTCAAGCGCCCGGCTCGCACCGGCCGCAACCC
GCGCACTGGCGAGCAGATTGACATTCCGGCTTCCTACGGCGTTCGTATCTCCGCTGGCTC
CCTGCTGAAGAAGGCCGTCACCGAGTATGGACGGAAGAAGCGCAGGCAGCGACGGCGAT
GATCTAGACTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCG
AGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGA
CTAGTTCAGCGAAGCTTCGACGAGAGCACATCGCCCGCCTCAGGCGACGAGAGCACATC
GCCCGCCTCGGTCGGCCGCGAGGCGACGAGAGCGCCTCGAAGCTTTTTTTTTTTTGGGGC
GGCTTTTTTTTTTTAAGCTTTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGC
GCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGC
TTAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTAGGTA
CTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAA
ATACCGCACAGAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATAT
TTGAATGTATTTAGAAAGCGGCCGCGTTAGGCGTTTTCGCGATGTACGGGCCAGATATACG
CGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGC

CCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCA
ACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGAC
TTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGG
CATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGT
CATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTT
GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGC
CTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCT
CCGGACTCTAGAGGATCGAAGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCA
CCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGC
GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTG
CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCG
TGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCA
TGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA
CCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC
ATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGC
CACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCC
GCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCT
GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCG
CCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAGGAATTCTTCGATCCCTACCGGT
TAGTAATGAGTTTAAACGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGG
AACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGGTGTTGGGTCGTTTG
TTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCC
ATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTG
AAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCGCGGCCGCT
CGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCG
CAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTG
ACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAA
CGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATGGGATCCATCA
TGCCTCCTCTAGACCAGCCAGGACAGAAATGCCTCGACTTCGCTGCTACCCAAGGTTGCC
GGGTGACGCACACCGTGGAAACGGATGAAGGCACGAACCCAGTGGACATAAGCCTGTTC

GGTTCGTAAGCTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGAC
CGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTT
GGGGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATG
TTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTA
AACATCATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTATCAGAGGTAGTTGGC
GTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGTACGGCTCCGCAGTG
GATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTT
GATGAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGA
GAGAGCGAGATTCTCCGCGCTGTAGAAGTCACCATTGTTGTGCACGACGACATCATTCCGT
GGCGTTATCCAGCTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTG
CAGGTATCTTCGAGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAG
AGAACATAGCGTTGCCTTGGTAGGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGA
ACAGGATCTATTTGAGGCGCTAAATGAAACCTTAACGCTATGGAACTCGCCGCCCGACTGG
GCTGGCGATGAGCGAAATGTAGTGCTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACC
GGCAAAATCGCGCCGAAGGATGTCGCTGCCGACTGGGCAATGGAGCGCCTGCCGGCCCA
GTATCAGCCCGTCATACTTGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGATCGCTTG
GCCTCGCGCGCAGATCAGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAG
GTAGTCGGCAAATAACCCTCGAGCCACCCATGACCAAAATCCCTTAACGTGAGTTACGCGT
CGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTT
CTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGGATCCATGCAG
CTGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
GCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTT
ATCCACAGAACGTACGATGTGAGCAAAAGGCCAGCAAAAGGCCAGGGACCGTAAAAAGGC
CGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACG
CTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGG
AAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTT
CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGT
AGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC
GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT
TGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCT
GAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCT
GGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACGGGCGTACGTCTTCCTTTTTCATCCCGGAGACGGTCACA

GCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGT
TGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTG
GCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATA
CCGCACAGAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACC
TGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGC
CCTTTCGTCTTCAAGAAAGATCTCCATGGGAGAACCACGGGCCGGACGGATACCAGCCGC
CCTCATACGAGCCGGTCAACCCCGAACGCAGGACCGCCCAGACGCCTTCCGATGGCCTG
ATCTGACGTCCGAAAAAAGGCGCCGTGCGCCCTTTTTAAATCTTTTAAAATCTTTTTACATT
CTTTTAGGCCCTCCGCAGCCCTACTCTCCCAACGGGTTTCGGACGGTACTTAGTACAAAAG
GGGAGCGAACTTAGTACAAAAGGGGAGCGAACTTAGTACAAAAGGGGAGCGAACTTAGTA
CAAAAGGGGAGCGAACTAGTAAATAAATAAACTTTGTACTAGTATGGAGTCATGTCCAATGA
GATCGTGAAGTTCAGCAACCAGTTCAACAACGTCGCGCTGAAGAAGTTCGACGCCGTGCA
CCTGGACGTGCTCATGGCGATCGCCTCAAGGGTGAGGGAGAAGGGCACGGCCACGGTGG
AGTTCTCGTTCGAGGAGCTGCGCGGCCTCATGCGATTGAGGAAGAACCTGACCAACAAGC
AGCTGGCAGACAAGATCGTGCAGACGAACGCGCGCCTGCTGGCGCTGAACTACATGTTCG
AGGATTCGGGCAAGATCATCCAGTTCGCGCTGTTCACGAAGTTCGTCACCGACCCGCAGG
AGGCGACTCTCGCGGTTGGGGTCAACGAGGAGTTCGCTTTCCTGCTCAACGACCTGACCA
GCCAGTTCACGCGCTTCGAGCTGGCCGAGTTCGCCGACCTCAAGAGCAAGTACGCCAAG
GAGTTCTACAGGCGCGCCAAGCAGTACCGCAGCTCGGGAATCTGGAAGATCAGCCGCGA
TGAGTTCTGCCGACTGCTTGGCGTATCCGATTCCACGGCAAAATCCACCGCCAACCTGAA
CAGGGTCGTGCTGAAGACGATCGCCGAAGAGTGTGGGCCTCTCCTTGGCCTGAAGATCGA
GCGCCAGTACGTGAAACGCAGGCTGTCGGGCTTCGTGTTCACGTTCGCCCGCGAGACCC
CTCCGGTGATCGACG

SEQ ID NO: 2: pFRG1.5 SHT

ACGCGCTGGAGATGTTCAACGAGTAGATCGCCACGGCGACCTCCTTCCACGCGTGCGGG
CACGGGGATTCTCAAGGGGCCGGCCCGAGGCCCCTTGAGCCCGCCGGGAGGCGCCCCC
GGCAGGGCGGGAATCCAAAGGGCGGAGCCCTGTGGCCCTCCCCGGGCAGGGGCGGGAT
CGTCAAGGGCGGAGCCCTTGGCCCCCTCGGGAGAGCGCACTGACACAATGCTACCTCCG
GTAGCATTAAGTGCGCCCTCCGCCATGCGGAGGGACGGGCCGCGACCGGATATGCGGGG
AACGTCCACGACGCGTCTTCCGTGTCCTCCGTCCTGCCTTGTGCCGTCGATTATAATCCTC
GGATAACGGACGATGATTGAACCGATTGGAGGAAACGAGATGCCGAAGAGTTTCGCGCAG
CAGATCGAGGACGACGAGAACAAGATCAAGCGCATCCGGGAGCACCAGCGCATGGTCAG
GGCCAAACAAGCCAAACAGGAGCGCAACGCCCGCACCAAAAGGCTCGTGGAGACCGGAG
CGATAGTGGAAAAAGCGCACGGCGGCGCGTACGACGACGAAGGACGGCAGACGTTCTCG
GATGGTCTGAACGGCATCATCTCGGTCTACGACCCGTCCCGCGGCGGCAACGTGGACAT
GAGAGTCATCGACGTAATCGACCGGCGCATCCCCAGATTGCCAAGGTCCGAAACCACAAC
AGGCACGGCAGCGGCGGCATCGCGAACCGTGCAAGCCACCGCACCACAGCCAGCCCAC
GCCCAACCGCAAAGCTTCACGCCCAACCCCCAGCGCATCGAACACCGGACAGGAGCACA
GCAGCCGGATCGGTGGGCCTAGCCGGGACGGCGCAGCCCTCGCTCCGCTGGGCTGAAA
TATCTGACTTGGTTCGTAACATATTTCCGACATGTCGGAAATATGTCGTATCATCGTTGCTA
TGAGTACTGAACTTCGTGAGCAGTGGGAGCAGCTTTACCTGCCGCTGCGCCCGCTGTGCA
CGAACGACTTCATCGAGGGCGTGTACCGGCAGCCCAGGGCGAAGGCGCTGGAGGGCTAC
CGGTACATCGAGGCGAACCCCAAGACCGTCAGCGACCTGCTCGTGGTCGACATCGACGA
CGCGAACGCCCGCGCGATGGCCCTATGGGAGCATGAGGGGATGCTGCCCAACGTGATCG
TGGAGAACCCGAGAAACGGCCACGCGCACGCCGTGTGGGCGCTGGCCGCCCCGTTCAGT
CGCACCGAATACGCGCGGCGCAAGCCGCTCGCCCTGGCCGCGGCGGTCACGGAGGGGC
TGCGCCGCTCGTGCGACGGGGACAAGGGATATTCGGGGCTGATGACCAAGAACCCGCTG
CACGAGGCGTGGAACAGCGAACTGGTCACCGATCACCTGTACGGGCTGGACGAACTGCG
CGAGGCGCTGGAGGAGTCCGGGGACATGCCGCCGGCCTCGTGGAAGCGCACGAAACGG
CGCAACGTGGTCGGATTGGGCCGCAACTGCCATCTGTTCGAGACGGCGCGCACATGGGC
GTACAGGGAGGTGCGCCACCACTGGGGAGACAGCGAGGGGCTGCGGCTGGCCATCTCG
GCCGAGGCGCACGAAATGAACGCCGAACTGTTCCCCGAGCCACTGGCATGGTCGGAGGT
CGAGCAGATCGCCAAGAGCATCCACAAGTGGATCGTCACCCAAAGCCGCATGTGGCGCG
ACGGCCCCACCGTCTACGAGGCGACATTCATCACCGTCCAAAGCGCCCGAGGGAAAAAGT
CAGGCAAGGCCAGACGAGAAACATTCGAACTATTCGCGAGCGAGGAGATGACGTAATGGT
TATTCAAACGTTTTTGGGGCGGCTTTTGGGCCCATGTGAGCAAAAGGCCAGCAAAAGGCC

AGGGACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAG
CATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC
CAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACC
GGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTA
GGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG
TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA
CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTT
GGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAG
AAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTTTTTTTGGGGCGGCTTTGA
ATTCTTTTTTTTGGGGCGGCTTTTTTTTTATGCGCTCACGCAACTGGTCCAGAACCTTGACC
GAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTG
GGGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTT
TGATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAAA
CATCATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTATCAGAGGTAGTTGGCGT
CATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGTACGGCTCCGCAGTGGA
TGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTTGAT
GAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAG
AGCGAGATTCTCCGCGCTGTAGAAGTCACCATTGTTGTGCACGACGACATCATTCCGTGG
CGTTATCCAGCTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGCAG
GTATCTTCGAGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGA
ACATAGCGTTGCCTTGGTAGGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACA
GGATCTATTTGAGGCGCTAAATGAAACCTTAACGCTATGGAACTCGCCGCCCGACTGGGC
TGGCGATGAGCGAAATGTAGTGCTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGG
CAAAATCGCGCCGAAGGATGTCGCTGCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGT
ATCAGCCCGTCATACTTGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGATCGCTTGGC
CTCGCGCGCAGATCAGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGT
AGTCGGCAAATAAGGTACCGTTAGGCGTTTTCGCGATGTACGGGCCAGATATACGCGTTG
ACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATA
TATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGA
CCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC
CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT
ATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTA

TGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG
CTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTC
ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAAT
CAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGG
CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGG
AGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGG
ACTCTAGAGGATCGAAGCTAGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGG
GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC
CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA
CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC
GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCA
CAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGC
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGTAGATCTTCGATCCCTACCGGTTAGTAATG
AGTTTAAACGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGC
GCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGGTGTTGGGTCGTTTGTTCATAAA
CGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGTCCCCATTGGGG
CCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCC
AGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCTTTTTGGGGCGGCTTTT
CTCGAGTCCTTCCTTAAGGACGTGCTCGTCAATTTTTGTTTTGAGAGTCATCTATTCGGATG
CTTTTCATGAAGTTTTTTATGACCCTGACCGGCACCCTGCGCAAGGCCTTCGCCACCACCC
TGGCCGCCGCCATGCTGATCGGCACCCTGGCCGGCTGCTCCTCCGCCGCATACAACAAG
TCTGACCTCGTTTCGAAGATCGCCCAGAAGTCCAACCTGACCAAGGCTCAGGCCGAGGCT
GCTGTTAACGCCTTCCAGGATGTGTTCGTCGAGGCTATGAAGTCCGGCGAAGGCCTGAAG
CTCACCGGCCTGTTCTCCGCTGAGCGCGTCAAGCGCCCGGCTCGCACCGGCCGCAACCC
GCGCACTGGCGAGCAGATTGACATTCCGGCTTCCTACGGCGTTCGTATCTCCGCTGGCTC
CCTGCTGAAGAAGGCCGTCACCGAGTATGGACGGAAGAAGCGCAGGCAGCGACGGCGAT
GAGGAAGAAGCGCAGGCAGCGACGGCGATGAGGGTTTGCGCTTGCGTCGTGGAGGGAG
CGGAACGCCGAAAAAGGATCC

SEQ ID NO: 3: Lac I DNA binding domain

AAATATGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTCATCAGACCGTTTCCC
GCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCG
ATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCG
TTGCTGATT

SEQ ID NO: 4: HU DNA binding domain

GCATACAACAAGTCTGACCTCGTTTCGAAGATCGCCCAGAAGTCCAACCTGACCAAGGCTC
AGGCCGAGGCTGCTGTTAACGCCTTCCAGGATGTGTTCGTCGAGGCTATGAAGTCCGGCG
AAGGCCTGAAGCTCACCGGCCTGTTCTCCGCTGAGCGCGTCAAGCGCCCGGCTCGCACC
GGCCGCAACCCGCGCACTGGCGAGCAGATTGACATTCCGGCTTCCTACGGCGTTCGTATC
TCCGCTGGCTCCCTGCTGAAGAAGGCCGTCACCGAG

SEQ ID NO: 5: Mer R DNA binding domain

GAAAACAATTTGGAGAACCTGACCATTGGCGTTTTCGCCAGGACGGCCGGGGTCAATGTG
GAGACCATCCGGTTCTATCAGCGCAAGGGCTTGCTCCCGGAACCGGACAAGCCTTACGGC
AGCATTCGCCGCTATGGCGAGACGGATGTAACGCGGGTGCGCTTCGTGAAATCAGCCCAG
CGGTTGGGCTTCAGCCTGGATGAGATCGCCGAGCTGCTGCGGCTGGAGGATGGCACCCA
TTGCGAGGAAGCCAGCAGCCTGGCCGAGCACAAGCTCAAGGACGTGCGCGAGAGGATGG
CTGACCTGGCGCGCATGGAGGCCGTGCTGTCTGATTTGGTGTGCGCCTGCCATGCGCGA
AGGGGGAACGTTTCCTGCCCGCTGATCGCGTCACTACAGGGTGGAGCAAGCTTGGCAGG
TTCGGCTATGCCT

SEQ ID NO: 6: Zinc finger DNA binding domain.

GAAAAACTGCGCAACGGCAGCGGCGATCCGGGCAAAAAAAAACAGCATGCGTGCCCGGA
ATGCGGCAAAAGCTTTAGCCAGAGCAGCGATCTGCAGCGCCATCAGCGCACCCATACCGG
CGAAAAACCGTATAAATGCCCGGAATGCGGCAAAAGCTTTAGCCGCAGCGATGAACTGCA
GCGCCATCAGCGCACCCATACCGGCGAAAAACCGTATAAATGCCCGGAATGCGGCAAAAG
CTTTAGCCGCAGCGATCATCTGAGCCGCCATCAGCGCACCCATCAGAACAAAAAA

SEQ ID NO: 7: Coding sequence for SMT protein

ATGACCCTGACCGGCACCCTGCGCAAAGCGTTTGCGACCACCCTGGCGGCGGCGATGCT
GATTGGCACCCTGGCGGGCTGCAGCAGCGCGGAAAACAATTTGGAGAACCTGACCATTGG

CGTTTTCGCCAGGACGGCCGGGGTCAATGTGGAGACCATCCGGTTCTATCAGCGCAAGG
GCTTGCTCCCGGAACCGGACAAGCCTTACGGCAGCATTCGCCGCTATGGCGAGACGGAT
GTAACGCGGGTGCGCTTCGTGAAATCAGCCCAGCGGTTGGGCTTCAGCCTGGATGAGATC
GCCGAGCTGCTGCGGCTGGAGGATGGCACCCATTGCGAGGAAGCCAGCAGCCTGGCCGA
GCACAAGCTCAAGGACGTGCGCGAGAGGATGGCTGACCTGGCGCGCATGGAGGCCGTGC
TGTCTGATTTGGTGTGCGCCTGCCATGCGCGAAGGGGGAACGTTTCCTGCCCGCTGATCG
CGTCACTACAGGGTGGAGCAAGCTTGGCAGGTTCGGCTATGCCTTATGGACGGAAGAAGC
GCAGGCAGCGACGGCGATGA

SEQ ID NO: 8: Coding sequence for SHT protein

ATGACCCTGACCGGCACCCTGCGCAAAGCGTTTGCGACCACCCTGGCGGCGGCGATGCT
GATTGGCACCCTGGCGGGCTGCAGCAGCGCGGCATACAACAAGTCTGACCTCGTTTCGAA
GATCGCCCAGAAGTCCAACCTGACCAAGGCTCAGGCCGAGGCTGCTGTTAACGCCTTCCA
GGATGTGTTCGTCGAGGCTATGAAGTCCGGCGAAGGCCTGAAGCTCACCGGCCTGTTCTC
CGCTGAGCGCGTCAAGCGCCCGGCTCGCACCGGCCGCAACCCGCGCACTGGCGAGCAG
ATTGACATTCCGGCTTCCTACGGCGTTCGTATCTCCGCTGGCTCCCTGCTGAAGAAGGCC
GTCACCGAGTATGGACGGAAGAAGCGCAGGCAGCGACGGCGATGA

SEQ ID NO: 9: Coding sequence for SLT

ATGACCCTGACCGGCACCCTGCGCAAAGCGTTTGCGACCACCCTGGCGGCGGCGATGCT
GATTGGCACCCTGGCGGGCTGCAGCAGCGCGAAATATGTAACGTTATACGATGTCGCAGA
GTATGCCGGTGTCTCTCATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTC
TGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCG
CGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTTATGGACGGAAGAAGCGCAG
GCAGCGACGGCGATGA

SEQ ID NO: 10: cODING SEQUENCE FOR SZT

ATGACCCTGACCGGCACCCTGCGCAAAGCGTTTGCGACCACCCTGGCGGCGGCGATGCT
GATTGGCACCCTGGCGGGCTGCAGCAGCGCGGAAAAACTGCGCAACGGCAGCGGCGATC
CGGGCAAAAAAAAACAGCATGCGTGCCCGGAATGCGGCAAAAGCTTTAGCCAGAGCAGCG
ATCTGCAGCGCCATCAGCGCACCCATACCGGCGAAAAACCGTATAAATGCCCGGAATGCG
GCAAAAGCTTTAGCCGCAGCGATGAACTGCAGCGCCATCAGCGCACCCATACCGGCGAAA
AACCGTATAAATGCCCGGAATGCGGCAAAAGCTTTAGCCGCAGCGATCATCTGAGCCGCC

ATCAGCGCACCCATCAGAACAAAAAATATGGACGGAAGAAGCGCAGGCAGCGACGGCGAT
GA

SEQ ID NO: 18: Alpha-amylase Sequence with extra amino acids that permit cleavage

Atgaaacatcggaaacccgcaccggcctggcataggctggggctgaagattagcaagaaagtggtggtcggcatcaccgccgcg
gcgaccgccttcggcggactggcaatcgccagcaccgcagcacaggccagcacc

SEQ ID NO: 19: 46 amino acid Alpha amylase Signal peptide with putative cleavage site

MKHRKPAPAWHRLGLKISKKVVVGITAAATAFGGLAIASTAAQAST

SEQ ID NO: 20: Cleaved alpha amylase signal peptide

atgaaacatcggaaacccgcaccggcctggcataggctggggctgaagattagcaagaaagtggtggtcg
gcatcaccgccgcggcgaccgccttcggcggactggcaatcgccagcaccgcagcacaggcc

SEQ ID NO: 21: 44 amino acid predicted cleaved alpha amylase signal peptide (no cleavage signal)

MKHRKPAPAWHRLGLKISKKVVVGITAAATAFGGLAIASTAAQA

SEQ ID NO: 22: Arabinosidase signal peptide coding sequence.

ACCCTGACCGGCACCCTGCGCAAAGCGTTTGCGACCACCCTGGCGGCGGCGATGCTGAT
TGGCACCCTGGCGGGCTGCAGCAGCGCG

SEQ ID NO: 23: Arabinosidase signal peptide

TLTGTLRKAFATTLAAAMLIGTLAGCSSA

SEQ ID NO: 24: SHT HYBRID PROTEIN (133 AMINO ACIDS)

MTLTGTLRKAFATTLAAAMLIGTLAGCSSAAYNKSDLVSKIAQKSNLTKAQAEAAVNAFQDVFV
EAMKSGEGLKLTGLFSAERVKRPARTGRNPRTGEQIDIPASYGVRISAGSLLKKAVTEYGRKKR
RQRRR

SEQ ID NO: 25: SLT HYBRID PROTEIN (104 AMINO ACIDS)

MTLTGTLRKAFATTLAAAMLIGTLAGCSSAKYVTLYDVAEYAGVSHQTVSRVVNQASHVSAKTR
EKVEAAMAELNYIPNRVAQQLAGKQSLLIYGRKKRRQRRR

SEQ ID NO: 26: SMT HYBRID PROTEIN (184 AMINO ACIDS)

MTLTGTLRKAFATTLAAAMLIGTLAGCSSAENNLENLTIGVFARTAGVNVETIRFYQRKGLLPEP
DKPYGSIRRYGETDVTRVRFVKSAQRLGFSLDEIAELLRLEDGTHCEEASSLAEHKLKDVRERM
ADLARMEAVLSDLVCACHARRGNVSCPLIASLQGGASLAGSAMPYGRKKRRQRRR

SEQ ID NO: 27: SZT HYBRID PROTEIN (139 AMINO ACIDS)

MTLTGTLRKAFATTLAAAMLIGTLAGCSSAEKLRNGSGDPGKKKQHACPECGKSFSQSSDLQR
HQRTHTGEKPYKCPECGKSFSRSDELQRHQRTHTGEKPYKCPECGKSFSRSDHLSRHQRTH
QNKKYGRKKRRQRRR

SEQ ID NO: 28: SEQUENCE COMPRISING PDOHJR ORI

ACGCGCTGGAGATGTTCAACGAGTAGATCGCCACGGCGACCTCCTTCCACGCGTGCGGG
CACGGGGATTCTCAAGGGGCCGGCCCGAGGCCCCTTGAGCCCGCCGGGAGGCGCCCCC
GGCAGGGCGGGAATCCAAAGGGCGGAGCCCTGTGGCCCTCCCCGGGCAGGGGCGGGAT
CGTCAAGGGCGGAGCCCTTGGCCCCCTCGGGAGAGCGCACTGACACAATGCTACCTCCG
GTAGCATTAAGTGCGCCCTCCGCCATGCGGAGGGACGGGCCGCGACCGGATATGCGGGG
AACGTCCACGACGCGTCTTCCGTGTCCTCCGTCCTGCCTTGTGCCGTCGATTATAATCCTC
GGATAACGGACGATGATTGAACCGATTGGAGGAAACGAGATGCCGAAGAGTTTCGCGCAG
CAGATCGAGGACGACGAGAACAAGATCAAGCGCATCCGGGAGCACCAGCGCATGGTCAG
GGCCAAACAAGCCAAACAGGAGCGCAACGCCCGCACCAAAAGGCTCGTGGAGACCGGAG
CGATAGTGGAAAAAGCGCACGGCGGCGCGTACGACGACGAAGGACGGCAGACGTTCTCG
GATGGTCTGAACGGCATCATCTCGGTCTACGACCCGTCCCGCGGCGGCAACGTGGACAT
GAGAGTCATCGACGTAATCGACCGGCGCATCCCCAGATTGCCAAGGTCCGAAACCACAAC
AGGCACGGCAGCGGCGGCATCGCGAACCGTGCAAGCCACCGCACCACAGCCAGCCCAC
GCCCAACCGCAAAGCTTCACGCCCAACCCCCAGCGCATCGAACACCGGACAGGAGCACA
GCAGCCGGATCGGTGGGCCTAGCCGGGACGGCGCAGCCCTCGCTCCGCTGGGCTGAAA
TATCTGACTTGGTTCGTAACATATTTCCGACATGTCGGAAATATGTCGTATCATCGTTGCTA
TGAGTACTGAACTTCGTGAGCAGTGGGAGCAGCTTTACCTGCCGCTGCGCCCGCTGTGCA
CGAACGACTTCATCGAGGGCGTGTACCGGCAGCCCAGGGCGAAGGCGCTGGAGGGCTAC
CGGTACATCGAGGCGAACCCCAAGACCGTCAGCGACCTGCTCGTGGTCGACATCGACGA

CGCGAACGCCCGCGCGATGGCCCTATGGGAGCATGAGGGGATGCTGCCCAACGTGATCG
TGGAGAACCCGAGAAACGGCCACGCGCACGCCGTGTGGGCGCTGGCCGCCCCGTTCAGT
CGCACCGAATACGCGCGGCGCAAGCCGCTCGCCCTGGCCGCGGCGGTCACGGAGGGGC
TGCGCCGCTCGTGCGACGGGGACAAGGGATATTCGGGGCTGATGACCAAGAACCCGCTG
CACGAGGCGTGGAACAGCGAACTGGTCACCGATCACCTGTACGGGCTGGACGAACTGCG
CGAGGCGCTGGAGGAGTCCGGGGACATGCCGCCGGCCTCGTGGAAGCGCACGAAACGG
CGCAACGTGGTCGGATTGGGCCGCAACTGCCATCTGTTCGAGACGGCGCGCACATGGGC
GTACAGGGAGGTGCGCCACCACTGGGGAGACAGCGAGGGGCTGCGGCTGGCCATCTCG
GCCGAGGCGCACGAAATGAACGCCGAACTGTTCCCCGAGCCACTGGCATGGTCGGAGGT
CGAGCAGATCGCCAAGAGCATCCACAAGTGGATCGTCACCCAAAGCCGCATGTGGCGCG
ACGGCCCCACCGTCTACGAGGCGACATTCATCACCGTCCAAAGCGCCCGAGGGAAAAAGT
CAGGCAAGGCCAGACGAGAAACATTCGAACTATTCGCGAGCGAGGAGATGACGTAA

SEQ ID NO: 29 : E. COLI/PUC ORI.

ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGGACCGTAAAAAGGCCGCGTTGCTGGCGTTT
TTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGG
CGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCA
AGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTA
ACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTT
TTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCT
TTTCTACGGGTTTTTTTGGGGCGGCTTT

SEQ ID NO: 30: PB44 ORI

CCCTCATACGAGCCGGTCAACCCCGAACGCAGGACCGCCCAGACGCCTTCCGATGGCCT
GATCTGACGTCCGAAAAAAGGCGCCGTGCGCCCTTTTTAAATCTTTTAAAATCTTTTTACAT
TCTTTTAGGCCCTCCGCAGCCCTACTCTCCCAACGGGTTTCGGACGGTACTTAGTACAAAA
GGGGAGCGAACTTAGTACAAAAGGGGAGCGAACTTAGTACAAAAGGGGAGCGAACTTAGT
ACAAAAGGGGAGCGAACTAGTAAATAAATAAACTTTGTACTAGTATGGAGTCATGTCCAATG

AGATCGTGAAGTTCAGCAACCAGTTCAACAACGTCGCGCTGAAGAAGTTCGACGCCGTGC
ACCTGGACGTGCTCATGGCGATCGCCTCAAGGGTGAGGGAGAAGGGCACGGCCACGGTG
GAGTTCTCGTTCGAGGAGCTGCGCGGCCTCATGCGATTGAGGAAGAACCTGACCAACAAG
CAGCTGGCAGACAAGATCGTGCAGACGAACGCGCGCCTGCTGGCGCTGAACTACATGTTC
GAGGATTCGGGCAAGATCATCCAGTTCGCGCTGTTCACGAAGTTCGTCACCGACCCGCAG
GAGGCGACTCTCGCGGTTGGGGTCAACGAGGAGTTCGCTTTCCTGCTCAACGACCTGACC
AGCCAGTTCACGCGCTTCGAGCTGGCCGAGTTCGCCGACCTCAAGAGCAAGTACGCCAAG
GAGTTCTACAGGCGCGCCAAGCAGTACCGCAGCTCGGGAATCTGGAAGATCAGCCGCGA
TGAGTTCTGCCGACTGCTTGGCGTATCCGATTCCACGGCAAAATCCACCGCCAACCTGAA
CAGGGTCGTGCTGAAGACGATCGCCGAAGAGTGTGGGCCTCTCCTTGGCCTGAAGATCGA
GCGCCAGTACGTGAAACGCAGGCTGTCGGGCTTCGTGTTCACGTTCGCCCGCGAGACCC
CTCCGGTGATCGACGCCAGGCCCGTGGAGGCGAGGAAGACGGACGGCGACGGCAAGGG
CCATTGGACGAGCGTGGCGGGGTATGGCGAGGTGTTCACGACCACGGAGCTGTTCGACG
TGACGGCCGCGCGTGACCACTTCGACGGCACCGTGGAGGCCGGGGAATGCCGTTTCTGC
GCGTTTGACGCGCGCAACCGCGAACATCATGCGCGGAACGCCGGAAGGTTGTTCTAGCG
GCCGTGTCCGCGCCTCTGGGGCGGTTGCGCCTGCCATGGAGATCT

SEQ ID NO: 31: TAT

TATGGACGGAAGAAGCGCAGGCAGCGACGGCGA

SEQ ID NO: 32: P-beta MPG (gp41-SV40)

GGCGCGCTGTTTCTGGGCTTTCTGGGCGCGGCGGGCAGCACCATGGGCGCGTGGAGCC
AGCCGAAAAAAAAACGCAAAGTG

SEQ ID NO: 33: Transportan (Galanin-mastoparan)

GGCTGGACCCTGAACAGCGCGGGCTATCTGCTGGGCAAAATTAACCTGAAAGCGCTGGC
GGCGCTGGCGAAAAAAATTCTG

SEQ ID NO: 34: Pep-1 (Trp-rich motif-SV40)

AAAGAAACCTGGTGGGAAACCTGGTGGACCGAATGGAGCCAGCCGAAAAAAAAACGCCGC
GTG

SEQ ID NO: 35: GFP forward primer

GAATTCGCCACCATGGTGAGCAAGGG

SEQ ID NO: 36 GFP Reverse Primer (Compliment):

TGGACGAGCTGTACAAGTAGGAATTC

SEQ ID NO: 37: Spectinomycin Forward Primer:

ATGCGCTCACGCAACTGGTCCAGAACCTT

SEQ ID NO: 38: Spectinomycin Reverse Primer (Compliment):

TTATTTGCCGACTACCTTGGTGATCTCGC

SEQ ID NO: 39 CMV Forward Primer:

GTTAGGCGTTTTCGCGATGTACGGGCCAGATA

SEQ ID NO: 40: CMV Reverse Primer (Compliment):

GTGGCGAATTCTTCGATCCTCTAGAGTCCGGAG

SEQ ID NO: 41: MerR DNA recognition site

AAGCTTGTCGACCTGCAGCCTTGGCGCTTGACTCCGTACATGAGTACGGAAGTAAGGTTA
CGCTATCCTTGGCTGCAGAAGCTT

SEQ ID NO: 42: LacI repressor DNA binding site

AAGCTTCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCT
TTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGCTCACAAAAGCTT

SEQ ID NO: 43: Synthetic zinc finger DNA recognition site

AAGCTTTTTTTTTTTTTGGGGCGGCTTTTTTTTTTTTAAGCTT

SEQ ID NO: 44: HU DNA binding domain

AYNKSDLVSKIAQKSNLTKAQAEAAVNAFQDVFVEAMKSGEGLKLTGLFSAERVKRPARTGRN
PRTGEQIDIPASYGVRISAGSLLKKAVTE

SEQ ID NO: 45: Mer R DNA binding domain

MENNLENLTIGVFARTAGVNVETIRFYQRKGLLPEPDKPYGSIRRYGETDVTRVRFVKSAQRLG
FSLDEIAELLRLEDGTHCEEASSLAEHKLKDVRERMADLARMEAVLSDLVCACHARRGNVSCP
LIASLQGGASLAGSAMP

SEQ ID NO: 46: Zn finger DNA binding domain

MEKLRNGSGDPGKKKQHACPECGKSFSQSSDLQRHQRTHTGEKPYKCPECGKSFSRSDELQ
RHQRTHTGEKPYKCPECGKSFSRSDHLSRHQRTHQNKK

SEQ ID NO: 47: Lac I DNA binding domain

MKYVTLYDVAEYAGVSHQTVSRVVNQASHVSAKTREKVEAAMAELNYIPNRVAQQLAGKQSLL
I

SEQ ID NO: 48: Prokaryotic expression cassette comprising Hu Promoter and terminator

TCCTTCCTTAAGGACGTGCTCGTCAATTTTTGTTTTGAGAGTCATCTATTCGGATGCTTTTC
ATGAAGTTTTTTGGAAGAAGCGCAGGCAGCGACGGCGATGAGGGTTTGCGCTTGCGTCG

SEQ ID NO: 49: Eukaryotic Expression Cassette comprising CMV promoter, Kozak sequence and TK Poly A site and terminator, flanking GFP coding sequences to be expressed

GGTTAGGCGTTTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTAC
ATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTC
AATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTG
GAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGC
CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTT
ATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGC
GGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT
CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAA
TGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC

TATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTT
TTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGAGGATCGAAGCT
AGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGAC
CACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGC
ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCG
ATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGC
TGTACAAGTAGATCTTCGATCCCTACCGGTTAGTAATGAGTTTAAACGGGGGAGGCTAACT
GAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAG
AATAAAACGCACGGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTG
GCACTCTGTCGATACCCCACCGAGTCCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTT
TCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGC
GGCAGGCCCTGCCATAGCTTTTTGGGGCGGCTTTT

The embodiments and examples presented herein are illustrative of the general nature of the subject matter claimed and are not limiting. It will be understood by those skilled in the art how these embodiments can be readily modified and/or adapted for various applications and in various ways without departing from the spirit and scope of the subject matter claimed. Phrases, words and terms employed herein are illustrative and are not limiting. Where permissible by law, all references cited herein are incorporated by reference in their entirety. It will be appreciated that any aspects of the different embodiments disclosed herein may be combined in a range of possible alternative embodiments, and alternative combinations of features, all of which varied combinations of features are to be understood to form a part of the subject matter claimed. Particular embodiments may alternatively comprise or consist of or exclude any one or more of the elements disclosed.

THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE RIGHT OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A hybrid protein comprising:

at least one signal sequence;

5 at least one DNA binding domain; and

at least one cell penetrating peptide (CPP) domain.

2. The hybrid protein according to claim 1 wherein the at least one CPP domain is a TAT domain, a VP22 domain, an Antp domain, a Rev domain, a P-beta (gp41-SV40) domain, a Transportan (Galanin-mastoparan) domain or a Pep-1 (Trp-rich motif SV40) domain.

10 3. The hybrid protein according to claim 1 wherein the at least one secretion signal sequence is selected from the group consisting of an alpha amylase signal sequence, a truncated alpha amylase signal sequence and an alpha arabinosidase signal sequences.

4. The hybrid protein according to claim 1 wherein the at least one DNA binding domain is a sequence specific DNA binding domain.

15 5. The hybrid protein according to claim 1 wherein the at least one DNA binding domain comprises at least one domain selected from the group consisting of: a Zinc finger DNA binding domain, a homeobox DNA binding domain, a MerR DNA binding domain, and a HU DNA binding domain.

6. The hybrid protein according to claim 1 comprising an alpha arabinosidase signal
20 sequence, a TAT domain and a HU DNA binding domain.

7. A method for transforming a target cell with a desired DNA, the method comprising the step of:

contacting the target cell with a protein-DNA complex comprising the desired DNA and the hybrid protein according to claim 1.

25 8. The method according to claim 7 wherein the DNA is a plasmid.

9. The method according to claim 8 wherein the plasmid comprises an expression cassette for expressing the hybrid protein in a gram positive bacterium.

10. The method according to claim 8 wherein the at least one DNA binding domain is a sequence specific DNA binding domain and wherein the plasmid comprises at least one nucleotide sequence for binding to said at least one sequence specific DNA binding domain.

11. The method according to claim 8 wherein said at least one CPP sequence has at least 90% amino acid sequence identity over a sequence of at least 10 contiguous amino acids to SEQ ID NO: 11, 12, 13, 14, 15, 16, or 17.

12. The method according to claim 8 wherein said at least one signal sequence has at least 90% amino acid sequence identity over a sequence of at least 10 contiguous amino acids to SEQ ID NO: 19, 21 or 23.

13. The method according to claim 8 wherein said at least one DNA binding domain has at least 90% amino acid sequence identity over at least 10 amino acids to SEQ ID NO: 44, 45 or 46.

14. The method of claim 7, wherein said cell is a gram positive bacterium.

15. The method of claim 7, wherein said cell is a mammalian cell.

16. The method of claim 7 further comprising the step of contacting said DNA with said hybrid protein to form said protein-DNA complex.

17. A cell transformed with an exogenous DNA using the hybrid protein according to claim 1, wherein the cell is a gram positive bacterium or is a mammalian cell.

18. The use of the hybrid protein according to claim 1 to transform a gram positive bacterium or a mammalian cell.

19. The use of the hybrid protein according to claim 1 to transform a cell selected from the group consisting of: Staphylococcus, streptococcus, Bifidobacterium, lactococcus, lactobacillus, clostridium.

20. A kit for transforming a target cell with a desired DNA, the kit comprising a quantity of the hybrid protein according to claim 1, and instructions to:

form a complex of the said hybrid protein with the desired DNA; and

contact the said complex with said target cell.

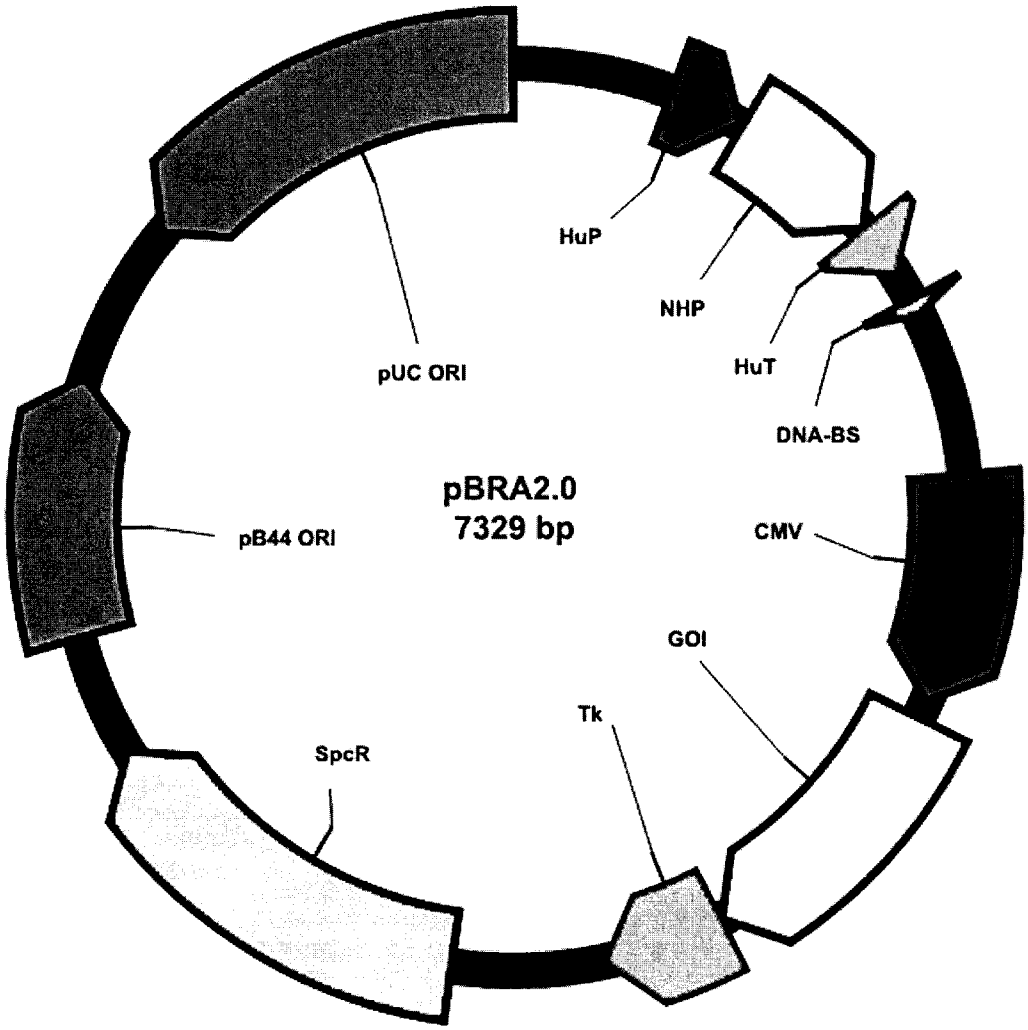
21. A gram positive bacterium able to synthesise the novel hybrid protein according to claim 1.

22. A hybrid protein comprising:

at least one CPP domain selected from the group consisting of a TAT domain, a VP22 domain, an Antp domain, a Rev domain, a P-beta (gp41-SV40) domain, a Transportan (Galanin-mastoparan) domain or a Pep-1 (Trp-rich motif SV40) domain;

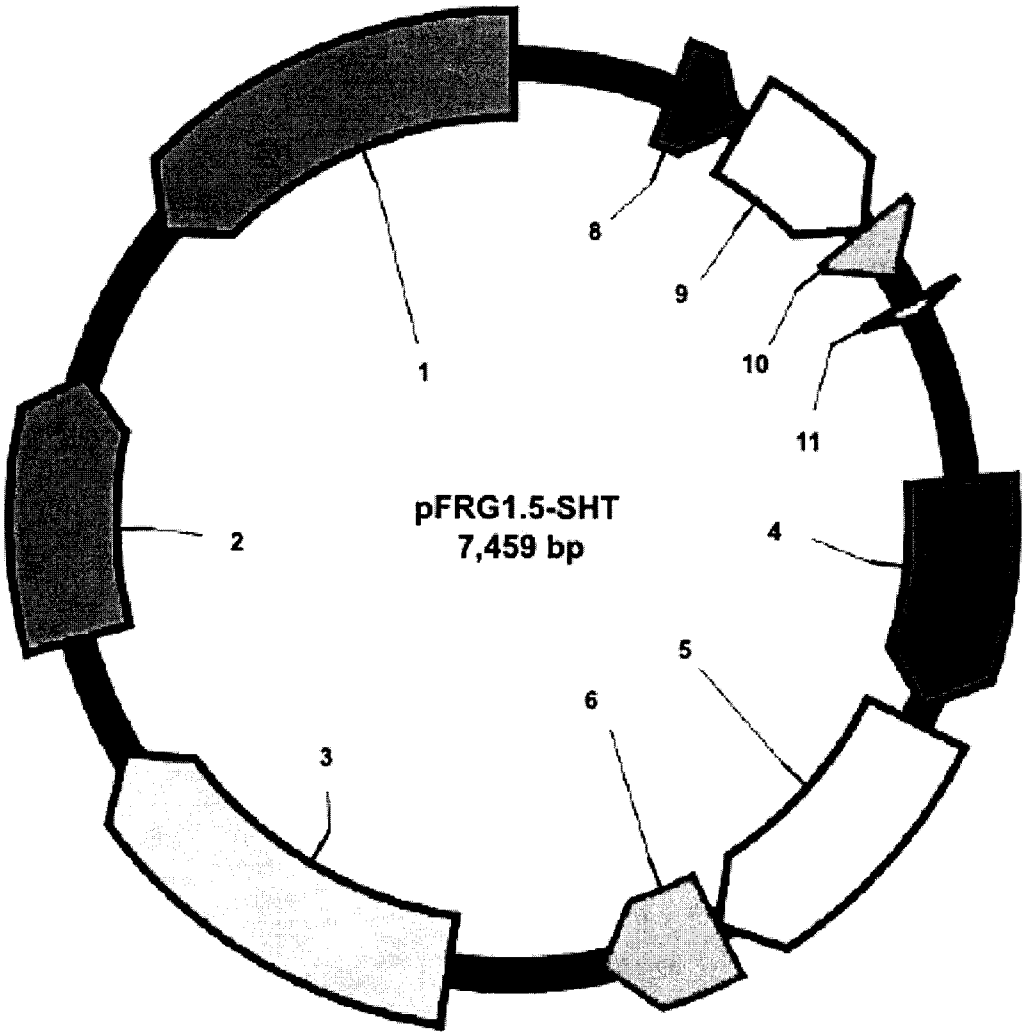
at least one DNA binding domain selected from the group consisting of a Zinc finger DNA binding domain, a homeobox DNA binding domain, a MerR DNA binding domain, and a HU DNA binding domain; and

at least one signal sequence selected from the group consisting of alpha amylase signal sequence, a truncated alpha amylase signal sequence and an alpha arabinosidase signal sequence.



pBRA 2.0-SHT

FIG. 1



pFRG 1.5 SHT

FIG. 2

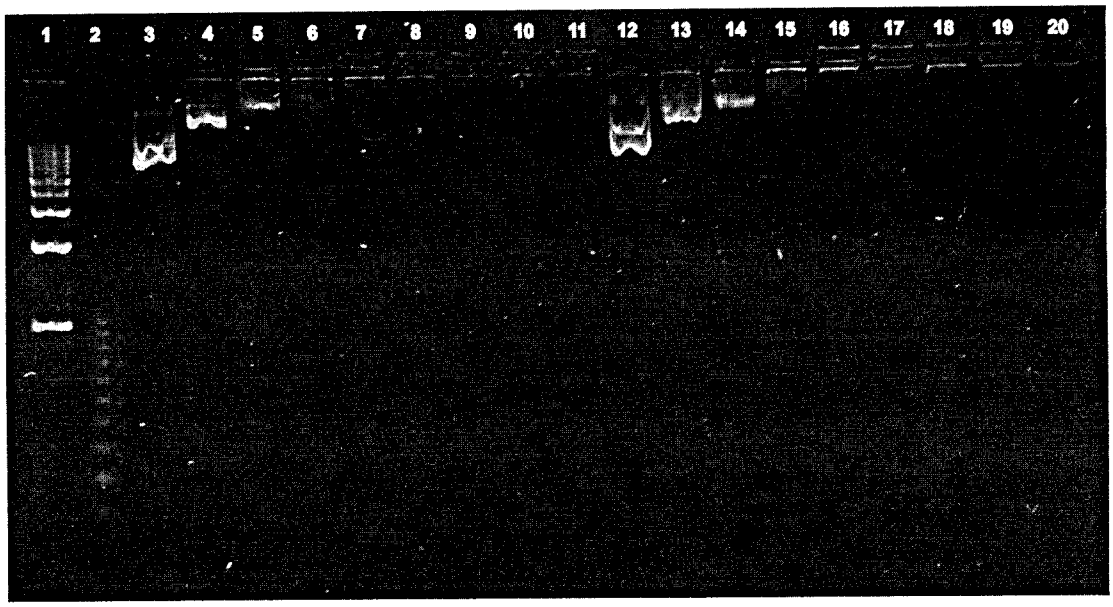


Figure 3. SHT and SZT interact with pBRA2.0 and result in gel-shift with increasing concentrations of peptide

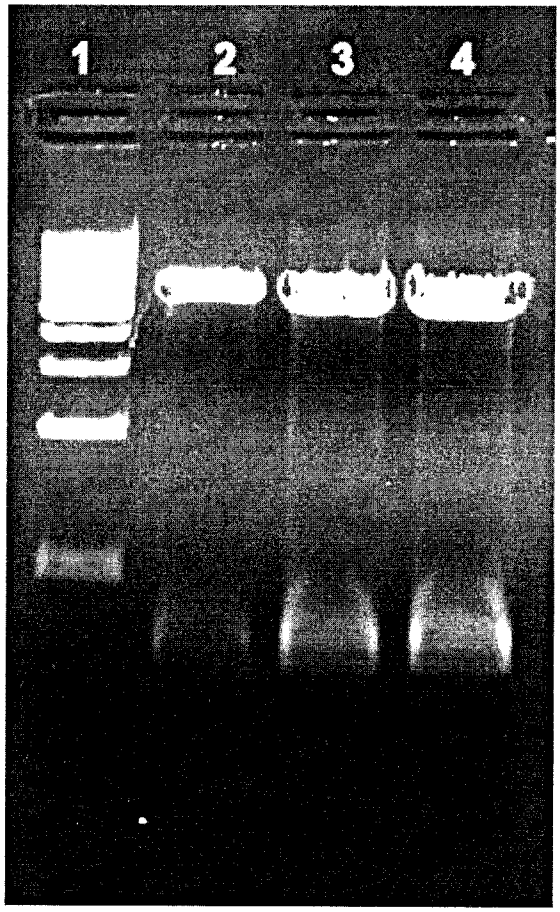


FIG. 4

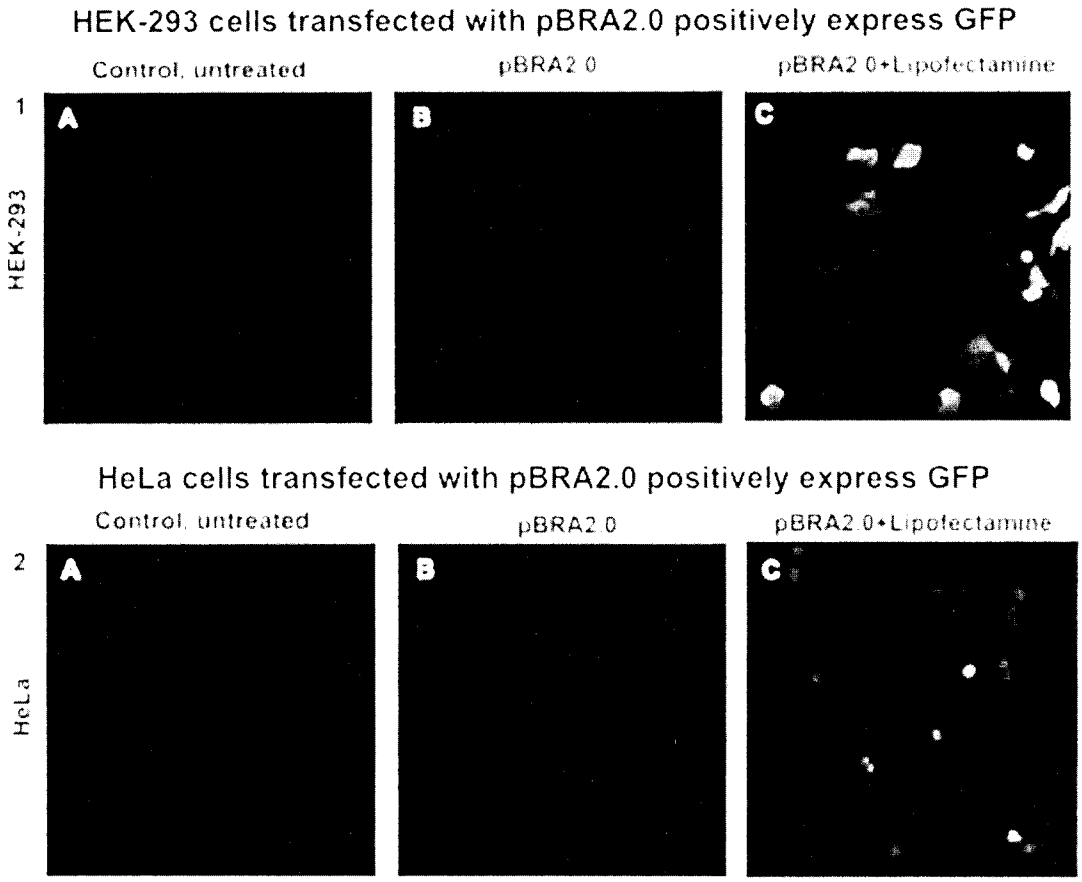

HEK-293 cells transfected with pBRA2.0 positively express GFP
Control, untreated
pBRA2.0
pBRA2.0+Lipofectamine
1
HEK-293
A
B
C
HeLa cells transfected with pBRA2.0 positively express GFP
Control, untreated
pBRA2.0
pBRA2.0+Lipofectamine
2
HeLa
A
B
C

FIG. 5

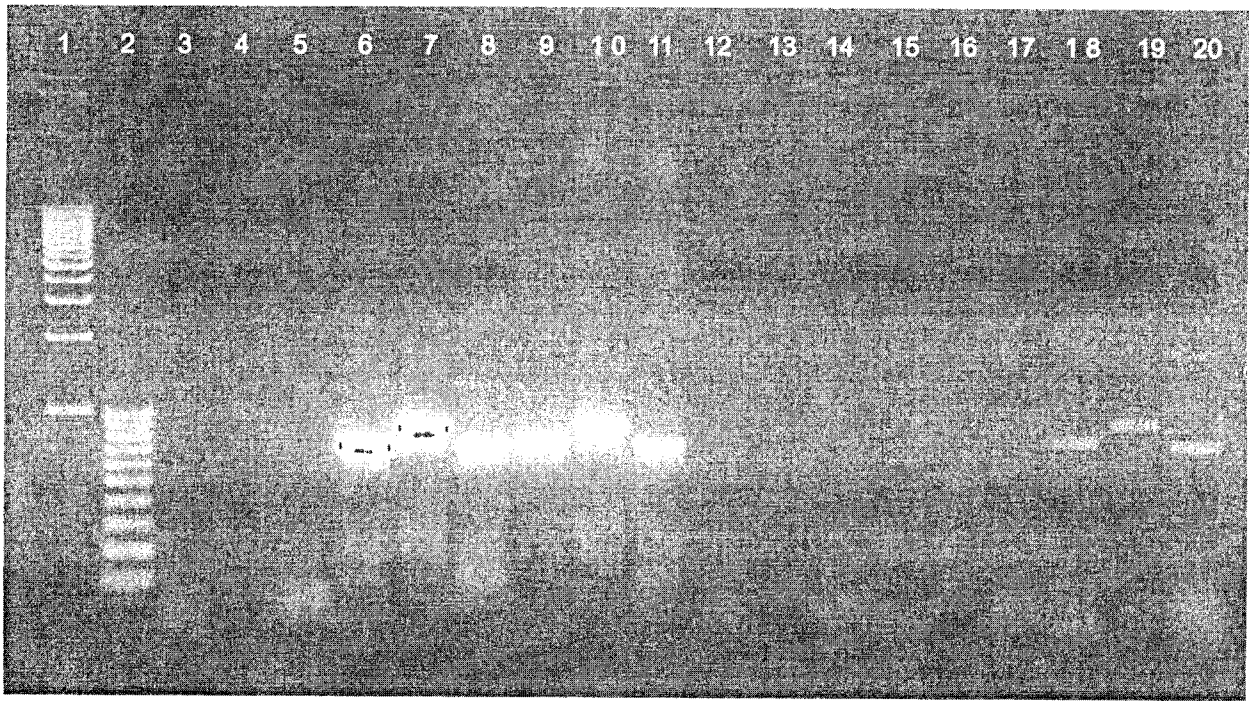


FIG. 6

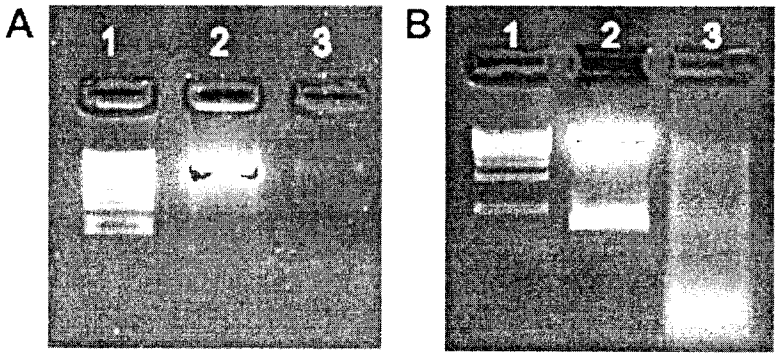


FIG. 7

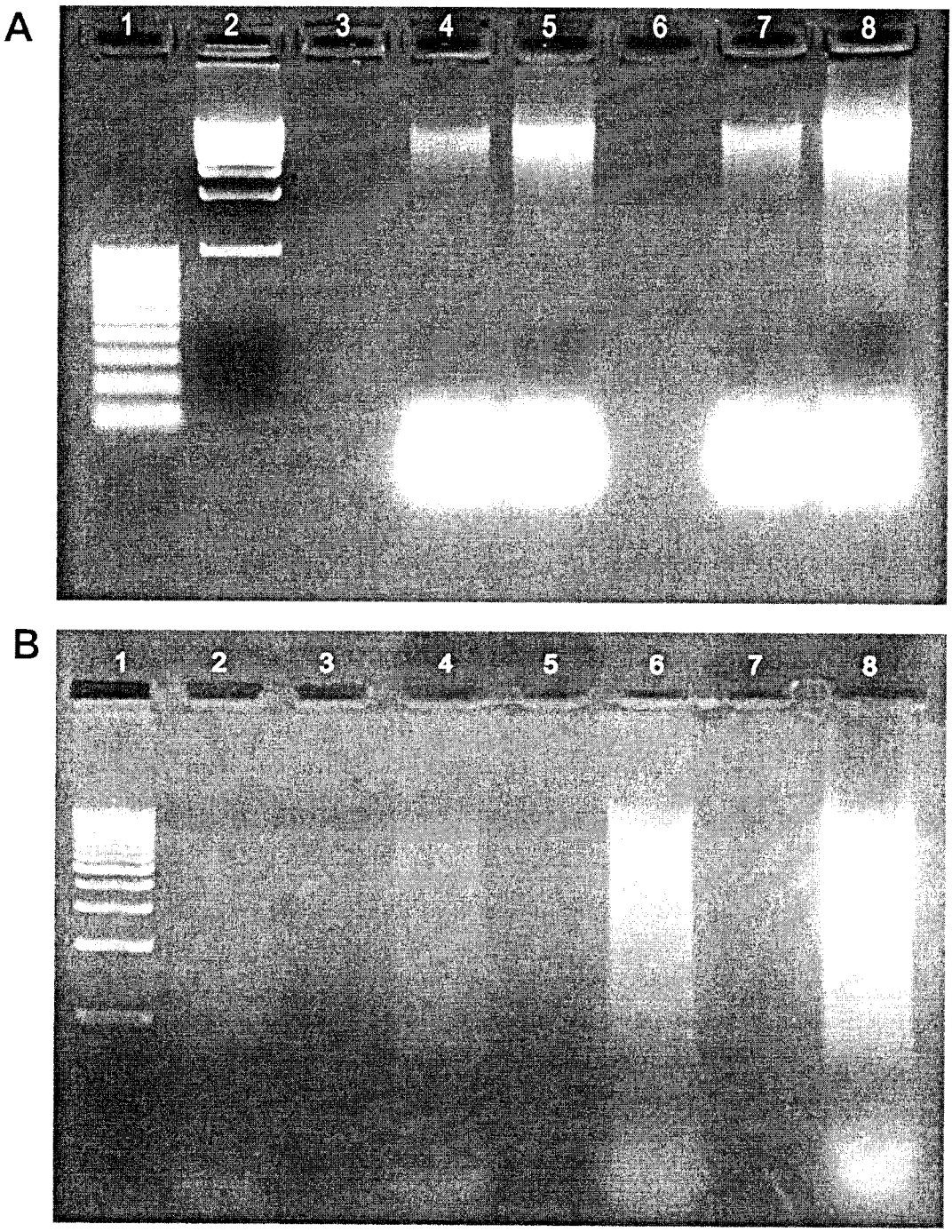


FIG. 8

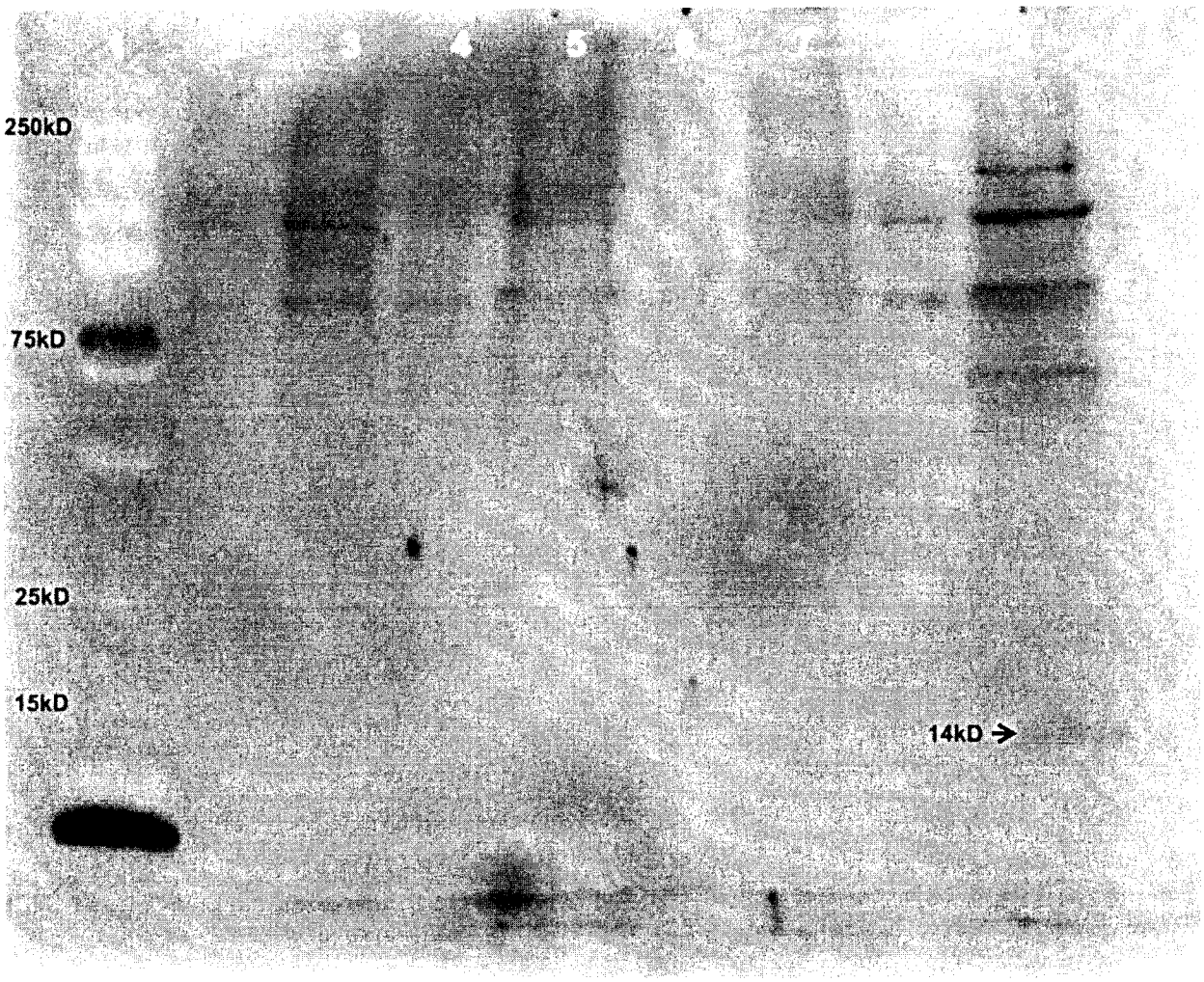


FIG. 9

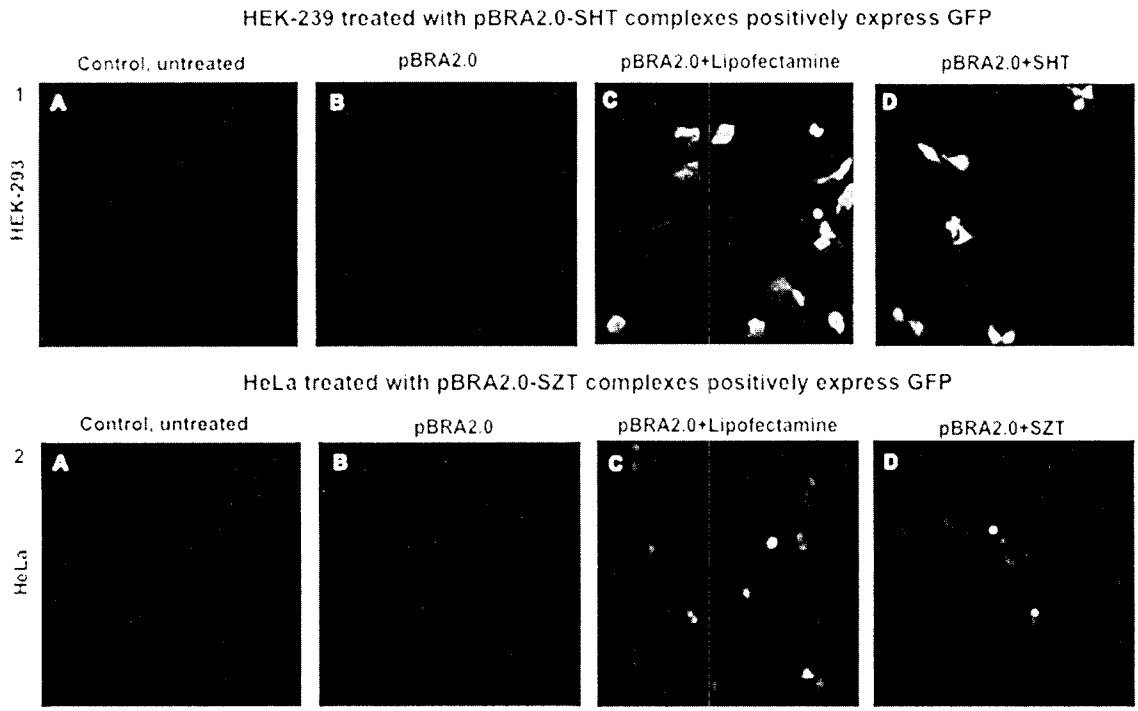


Figure 10 Therapeutic molecules pBRA2.0-SHT and pBRA2.0-SZT complexes can transfect and express GFP in mammalian cell lines.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2015/000084

A. CLASSIFICATION OF SUBJECT MATTER
IPC (2006.01): ***C07K 19/00***, ***C07K 14/025***, ***C07K 14/16***, ***C07K 14/435***, ***C12N 1/21***, ***C12N 15/62***, ***C12N 15/63***, ***C12N 15/85***, ***C12N 15/87***, ***C12N 5/10***, ***C12P 21/02***

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC (2006.01): C07K, C12N, C12P

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

Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)
Scopus, Total Patent

Keywords: hybrid protein, cell penetrating peptide, CPP, DNA binding domain, transform, bacteri*, delivery, transport, alpha amylase, bifidobacteri*

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	SMITH, J. et al. (2003) Novel "Three-in-One" Peptide Device for Genetic Drug Delivery. PROTEINS AND PEPTIDE LETTERS **10**: 1-7. ISSN: 1875-5305 *whole document*	1, 7, 15-18 and 20 2-6, 8-14, 19, 21 and 22
X Y	SERA, T. (2010) Generation of Cell-Permeable Artificial Zinc Finger Protein Variants. METHODS IN MOLECULAR BIOLOGY **649**: 91-96. ISSN: 1064-3745 *whole document*	1, 2, 4, 5, 7-11, 13, 15-18 and 20 3, 6, 12, 14, 19, 21 and 22
Y	KHOKHLOVA, E. V. et al. (October 2011) *Bifidobacterium longum* Modified Recombinant HU Protein as a Vector for Nonviral Delivery of DNA to HEK293 Human Cell Culture. BULLETIN OF EXPERIMENTAL BIOLOGY AND MEDICINE **151**: 663-667. ISSN: 1573-8221 *whole document*	1-22

☑ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"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
"A"	document defining the general state of the art which is not considered to be of particular relevance	"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
"E"	earlier application or patent but published on or after the international filing date	"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
"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)	"&"	document member of the same patent family
"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		

Date of the actual completion of the international search	Date of mailing of the international search report
10 April 2015 (10-04-2015)	04 May 2015 (04-05-2015)
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476	Authorized officer Kerry Ferguson (819) 994-9451

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GAO, S. et al. (March 2009) Bifunctional chimeric fusion proteins engineered for DNA delivery: Optimization of the protein to DNA ratio. BIOCHIM BIOPHYS ACTA **1970**: 198-207. ISSN: 0006-3002 *whole document*	1-22
Y	DENG, Q. et al. (September 2009) Signal peptide of Arabinosidase enhances secretion of interferon-alpha2b protein by Bifidobacterium longum. ARCH MICROBIOL **191**: 681-686. ISSN: 1432-072X *abstract*	1, 3, 6-9, 12, 14 and 17-22
A	BECHARA, C. et al. (2013) Cell-penetrating peptides: 20 years later, where do we stand? FEBS LETTERS **587**: 1693-1702. ISSN: 0014-5793 *whole document*	1-22
A	JOLIOT, A. et al. (March 2004) Transduction peptides: from technology to physiology. NATURE CELL BIOLOGY **6**: 189-196. ISSN: 1476-4679 *whole document*	1-22
A	MARGUS, H. et al. (March 2012) Cell-penetrating Peptides as Versatile Vehicles for Oligonucleotide Delivery. MOLECULAR THERAPY **20**: 525-533. ISSN: 1525-0024 *whole document*	1-22

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of the 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. ☑ Claim Nos.: 7-13, 15 and 16
because they relate to subject matter not required to be searched by this Authority, namely:

Claims 7-13, 15 and 16 are directed to a method for treatment of the human or animal body by surgery or therapy, which the International Searching Authority is not required to search under Rule 39.1(iv) of the PCT. However, this Authority has carried out a search based on the alleged use of the hybrid protein of claim 1 for transforming a target cell with a desired DNA.

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

3. ☐ Claim 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 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 claim 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 claim 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.