ENHANCED TRANSGENE EXPRESSION AND PROCESSING

FIELD OF THE INVENTION

The invention is directed at providing nucleic acid constructs and proteins that are involved in or act on metabolic pathways that mediate or influence cellular metabolism, e.g., translocation across the ER membrane and/or secretion across the cytoplasmic membrane as well as methods to influence cellular metabolism. The invention is also directed at the production and use of recombiant mammalian cells in which, e.g., translocation/secretion of a wide variety of heterologous proteins (transgene expression products) is altered. The methods, nucleic acid constructs are generally designed to improve transgene expression.

BACKGROUND OF THE INVENTION

The biotechnological production of therapeutical proteins as well as gene and cell therapy depends on the successful expression of transgenes introduced into a eukaryotic cell. Successful transgene expression generally requires integration of the transgene into the host chromosome and is limited, among others, by the number of transgene copies integrated and by epigenetic effects that can cause low or unstable transcription and/or high clonal variability. Failing or reduced transport of the transgene expression product out of the cell also often limits production of therapeutical proteins as well as gene and cell therapy.

The publications and other materials, including patents, patent applications and accession numbers, used herein to illustrate the invention and, in particular, to provide additional details respecting the practice are incorporated herein by reference in their entirety.

To increase and stabilize transgene expression in mammalian cells, epigenetic regulators are being increasingly used to protect transgenes from negative position effects (Bell and Felsenfeld, 1999). These epigenetic regulators include boundary or insulator elements, locus control regions (LCRs), stabilizing and antirepressor (STAR) elements, ubiquitously acting chromatin opening (UCOE) elements and the aforementioned matrix attachment regions (MARs). All of these epigenetic regulators have been used for recombinant protein production in mammalian cell lines (Zahn-Zabal et al., 2001; Kim et al., 2004) and for gene therapies (Agarwal et al., 1998; Allen et al., 1996; Castilla et al., 1998).

The transgene expression product often encounters different bottlenecks during processing and transport out of the cell: The cell that is only equipped with the machinery to process and transport its innate proteins can get readily overburdened by the transport of certain types of transgene expression products, especially when they are produced at abnormally high levels as often desired, letting the product aggregate within the cell and/or, e.g., preventing proper folding of a functional protein product.

Different approaches have been pursued to overcome transportation and processing bottlenecks. For example, CHO cells with improved secretion properties were engineered by the expression of the SM proteins Munc18c or Sly1, which act as regulators of membranous vesicles trafficking and hence secreted protein exocytosis (U.S. Patent Publication 20090247609). The X-box-binding protein 1 (Xbp1), a transcription factor that regulates secretory cell differentiation and ER maintenance and expansion, or various protein disulfide isomerases (PDI), have been used to decrease ER stress and increase protein secretion (Mohan et al. 2007). Other attempts to increase protein secretion included the expression of the chaperones ERp57, calnexin, calreticulin and BiP1 in CHO cells (Chung et al., 2004). Expression of a cold shock-induced protein, in particular the cold-inducible RNA-binding protein (CIRP), was shown to increase the yield of recombinant γ-interferon. Attempts were also made to overexpress proteins of the secretory complexes. However, for instance, Lakkaraju et al. (2008) reported that exogenous SRP14 expression in WT human cells (e.g. in cells that were not engineered to express low SRP14 levels) did not improve secretion efficiency of the secreted alkaline phosphatase protein.

Thus, there is a need for efficient, reliable transgene expression, e.g., recombinant protein production and for gene therapy. There is also a need to successfully transport the transgene expression product outside the cell.

This and other needs in the art are addressed by embodiments of the present invention.

SUMMARY OF THE INVENTION

The invention is, in one embodiment directed at a recombinant nucleic acid molecule comprising:

(a) a 5' and a 3' transposon-specific inverted terminal repeat (ITR),

(b) at least one nucleic acid sequence encoding a transgene expression processing (TEP) protein or a TEP functional RNA, located between the 5' and 3' ITRs and which is under the control of a promoter, and

(c) optionally at least one transgene also located between the 5' and 3' ITRs and which is under the control of a transgene promoter, wherein said nucleic acid molecule is optionally part of a vector.

The recombinant nucleic acid molecule may comprise at least one epigenetic regulatory element, in particular at least one MAR (matrix attachment region) element.

The MAR element may be located between the 5' and a 3' ITRs.A transgene such as an antibiotic resistance gene or a gene encoding an immunoglobulin, optionally under the control of a further promoter, may be located between the 5' and a 3' ITR such as between the 5' ITR and the MAR.

The TEP protein or TEP functional RNA may be a protein or a functional RNA that is, directly or indirectly, involved in integration of nucleic acid sequences into a genome, processing or translation of the transgene RNA product or is involved in ER translocation, secretion, processing, folding, ER-Golgy-plasma membrane transport, glycosylation and/or another post-translational modification of proteins such as transgene expression products.

The TEP protein may be a protein of the protein secretion pathway, a protein of the DNA recombination or repair pathways, a protein processing or metabolic protein including chaperones such as BiP, or a combination thereof.

The TEP protein may be one or more of the following proteins of the protein secretion pathway: hSRP14, hSec61α1, hSec61β, hSec61γ, hSRP54, hSRP9, hSRPRα, hSRPβ, and hCANX.

The TEP protein may also correspond to one or more of the following amino acid sequences of proteins of the protein secretion pathway: hSRP14 having SEQ ID NO: 13, hSec61α1 having SEQ ID NO: 15, hSec61β having SEQ ID NO: 17, hSec61γ having SEQ ID NO: 19, hSRP54 having SEQ ID NO: 21, hSRP9 having SEQ ID NO: 23, hSRPRα having SEQ ID NO: 25, hSRPβ having SEQ ID NO: 27, and hCANX having SEQ ID NO: 29 and/or may correspond to amino acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

The TEP protein may be one or more of the following protein processing or metabolic proteins: hUCP4, hCMPSAT, rST6Gal1, hCOMSC, hT-Synthase, hP4HA1, hP4HB, hGILZ, hCyPB, hNRF2, hHK1, hPDI, hPIN1, hSEPW1, hCALR, hDDOST, hHSP40, hATP5A1, hSERCA2, hPDIA4, hHSC70 /HSPA8, hHYOU1, hCMP-SAS, hBeclin-1, hERdj3, CHO-AGE, hWip1, hRTP4, hREEP2, hDPM1 and hDRiP78.

The TEP protein may also correspond to one or more of the following amino acid sequences of protein processing or metabolic proteins: hUCP4 having SEQ ID NO: 31, hCMPSAT having SEQ ID NO: 33, rST6Gal1 having SEQ ID NO: 35, hCOMSC having SEQ ID NO: 37, hT-Synthase having SEQ ID NO: 39, hP4HA1 having SEQ ID NO: 41, hP4HB having SEQ ID NO: 43, hGILZ having SEQ ID NO: 45, hCyPB having SEQ ID NO: 47, hNRF2 having SEQ ID NO: 49, hHK1 having SEQ ID NO: 51, hPDI having SEQ ID NO: 53, hPIN1 having SEQ ID NO: 55, hSEPW1 having SEQ ID NO: 57, hCALR having SEQ ID NO: 59, hDDOST having SEQ ID NO: 62, hHSP40 having SEQ ID NO: 64, hATP5A1 having SEQ ID NO: 66, hSERCA2 having SEQ ID NO: 68, hPDIA4 having SEQ ID NO: 70, hHSC70 /HSPA8 having SEQ ID NO: 72, hHYOU1 having SEQ ID NO: 74, hCMP-SAS having SEQ ID NO: 76, hBeclin-1 having SEQ ID NO: 78, hERdj3 having SEQ ID NO: 80, CHO-AGE having SEQ ID NO: 82, hWip1 having SEQ ID NO: 84, hRTP4 having SEQ ID NO: 86, hREEP2 having SEQ ID NO: 88, hDPM1 having SEQ ID NO: 90 and hDRiP78 having SEQ ID NO: 92 and/or may correspond to amino acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

The TEP protein may be a chaperone, in particular, a BiP protein, more in particular, a modified drosophila BIP protein derivative (DroBiP) having 80%, 90%, 95% or 100% sequence identity with the SEQ ID NO.: 60.

The MAR element may be selected from SEQ ID NOs: 1 (MAR 1-68), 2 (MAR 1_6), 3 (MAR X_S29), 4 (MAR S4), 5 (chicken lysozyme MAR), or preferably is an engineered, in particular rearranged counterpart and/or has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with any one of SEQ ID NOs: 1 to 5 or with any one of SEQ ID NOs: 6 to 10.

The TEP functional RNA within said cell may comprise/consist of nucleic acid sequences encoding a functional RNA, preferably a miRNA or a shRNA, that interferes with the expression of at least one protein of a DNA recombination or repair pathway, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2, Cyclin D1, Ercc, MDC1, Bard1, Ligase 1, Mre11 and/or 53BP1.

The TEP functional RNA may also interfere with expression of ngenes having at least 80%, 90%, 95%, 98% or 100% sequence identity with Rad51 having SEQ ID NO: 93, Rad51B having SEQ ID NO: 94, Rad51C having SEQ ID NO: 95, Rad51D having SEQ ID NO: 96, Xrcc2 having SEQ ID NO: 99, Xrcc3 having SEQ ID NO: 100, Rad52 having SEQ ID NO: 97, Rad54 having SEQ ID NO: 98 , Brca1 having SEQ ID NO: 101, Brca2 having SEQ ID NO: 102, Cyclin D1 having SEQ ID NO: 103, Ercc1 having SEQ ID NO: 104, MDC1 having SEQ ID NO: 105, Bard1 having SEQ ID NO: 106, Ligase 1 having SEQ ID NO: 107, Mre11 having SEQ ID NO: 108 and/or 53BP1 having SEQ ID NO: 109.

The recombinant nucleic acid molecule may be at least 5000, 6000, 7000, 8000, 90000 or 10000 bps long.

The 5' and a 3' ITRs may be 5' and 3' ITRs of the Sleeping Beauty or preferably PiggyBac Transposon.

Upon a first transfection of one of the recombinant nucleic acid molecules and a second, subsequent, transfection of a further recombinant nucleic acid molecule containing a transgene into a mammalian cell, transgene integration and/or expression may be increased in said cell relative to a cell not subject to said first transfection.

The TEP coding sequence or TEP functional RNA mentioned herein may be part of a vector including an expression vector. The vector may comprise a singular MAR element, two or more MAR elements, wherein said element(s) may be located between the 5' and 3' ITRs.

E.g., the vector may comprise two MAR elements. A first MAR element may be positioned upstream of the TEP or TEP functional RNA and a second MAR element may be positioned downstream of the TEP or TEP functional RNA, wherein the first MAR element may comprise a MAR 1_6 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 2, in particular a rearranged MARs based on MAR 1-6, more in particular elements that have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO: 8 (MARs 1_6R2) and the second MAR element may comprise a MAR 1-68 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 1.

The vector may also comprise a singular MAR element. The singular MAR element may be positioned downstream of the TEP or TEP functional RNA, wherein the singular MAR element may be a MAR 1-68 or a MAR X-29 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NOs. 1 or 3, in particular a rearranged MAR based on MAR 1-68 or a MAR X-29, in particular an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NOs: 6, 7 or 10 (MARs 1_68R, 1_68R2 or X_29R3) or 9, and may preferably a MAR X-29 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 3.

The TEP or TEP functional RNA may be under the control of an EF1 alpha promoter and is optionally followed by a BGH polyA signal.

The vector may comprise promoter(s) and/or enhancer(s) or fusions thereof such as GAPDH, SV40p, CMV, CHO EF1 alpha, CHO Actb and/or CHO Hspa5, or engineered fusions thereof, such as CGAPDH.

The promoters which are part of the vector may be GAPDH having SEQ ID NO: 111, SV40p having SEQ ID NO: 114, CMVp having SEQ ID NO: 113, CHO Ef1 alpha having SEQ ID NO:112, CHO Actb having SEQ ID NO: 115, CHO Hspa5 having SEQ ID NO: 116 , and/or fusions thereof such as CGAPDH having SEQ ID NO: 11, or may have nucleic acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

The invention is also directed at a method for expressing a TEP or TEP functional RNA comprising:
providing a recombinant mammalian cell comprising a transgene, and the vector is an expression vector which expresses the TEP or TEP functional RNA, wherein the TEP or TEP functional RNA expressed via said vector optionally increases an expression of a transgene in said mammalian cell by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70%.

The vector may comprise a singular MAR X-29 element and/or a nucleic acid sequence having at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 3 and wherein, after more than 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 weeks of cultivation, the TEP or TEP functional RNA expressed via said vector may increase an expression of a gene of interest by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70%.

The invention is also directed at a recombinant mammalian cell comprising not more than 20, 15, 10 or 5 of the recombinant nucleic acid molecule, preferably integrated into the genome of the cell as single copies.

As noted above, the TEP protein may be one or more of the following proteins of the protein secretion pathway: hSRP14, hSec61α1, hSec61β, hSec61γ, hSRP54, hSRP9, hSRPRα, hSRPβ, and hCANX.

The TEP protein may also correspond to one or more of the following amino acid sequences of proteins of the protein secretion pathway: hSRP14 having SEQ ID NO: 13, hSec61α1 having SEQ ID NO: 15, hSec61β having SEQ ID NO: 17, hSec61γ having SEQ ID NO: 19, hSRP54 having SEQ ID NO: 21, hSRP9 having SEQ ID NO: 23, hSRPRα having SEQ ID NO: 25, hSRPβ having SEQ ID NO: 27, and hCANX having SEQ ID NO: 29 and/or may correspond to amino acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences

The TEP protein may be one or more of the following protein processing or metabolic proteins: hUCP4, hCMPSAT, rST6Gal1, hCOMSC, hT-Synthase, hP4HA1, hP4HB, hGILZ, hCyPB, hNRF2, hHK1, hPDI, hPIN1, hSEPW1, hCALR, hDDOST, hHSP40, hATP5A1, hSERCA2, hPDIA4, hHSC70 /HSPA8, hHYOU1, hCMP-SAS, hBeclin-1, hERdj3, CHO-AGE, hWip1, hRTP4, hREEP2, hDPM1 and hDRiP78.

The TEP protein may also correspond to one or more of the following amino acid sequences of protein processing or metabolic proteins: hUCP4 having SEQ ID NO: 31, hCMPSAT having SEQ ID NO: 33, rST6Gal1 having SEQ ID NO: 35, hCOMSC having SEQ ID NO: 37, hT-Synthase having SEQ ID NO: 39, hP4HA1 having SEQ ID NO: 41, hP4HB having SEQ ID NO: 43, hGILZ having SEQ ID NO: 45, hCyPB having SEQ ID NO: 47, hNRF2 having SEQ ID NO: 49, hHK1 having SEQ ID NO: 51, hPDI having SEQ ID NO: 53, hPIN1 having SEQ ID NO: 55, hSEPW1 having SEQ ID NO: 57, hCALR having SEQ ID NO: 59, hDDOST having SEQ ID NO: 62, hHSP40 having SEQ ID NO: 64, hATP5A1 having SEQ ID NO: 66, hSERCA2 having SEQ ID NO: 68,

hPDIA4 having SEQ ID NO: 70, hHSC70 /HSPA8 having SEQ ID NO: 72, hHYOU1 having SEQ ID NO: 74, hCMP-SAS having SEQ ID NO: 76, hBeclin-1 having SEQ ID NO: 78, hERdj3 having SEQ ID NO: 80, CHO-AGE having SEQ ID NO: 82, hWip1 having SEQ ID NO: 84, hRTP4 having SEQ ID NO: 86, hREEP2 having SEQ ID NO: 88, hDPM1 having SEQ ID NO: 90 and hDRiP78 having SEQ ID NO: 92 and/or may correspond to amino acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

The TEP protein may be a chaperone, in particular, a BiP protein, more in particular, a engineered drosophila BIP protein derivative (DroBiP) having 80%, 90%, 95% or100% sequence identity with the SEQ ID NO.: 60.

The TEP functional RNA within the recombinant mammalian cell may comprise/consist of nucleic acid sequence(s) encoding a functional RNA, preferably a miRNA or a shRNA, that interferes with the expression of at least one recombination protein, preferably an HR gene, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2, Cyclin D1, Ercc1, MDC1, Bard1, Ligase 1, Mre11 and/or 53BP1. The nucleic acids may have at least 80%, 90%, 95%, 98% or 100% sequence identity with Rad51 having SEQ ID NO: 93, Rad51B having SEQ ID NO: 94, Rad51C having SEQ ID NO: 95, Rad51D having SEQ ID NO: 96, Xrcc2 having SEQ ID NO: 99, Xrcc3 having SEQ ID NO: 100, Rad52 having SEQ ID NO: 97, Rad54 having SEQ ID NO: 98 , Brca1 having SEQ ID NO: 101, Brca2 having SEQ ID NO: 102, Cyclin D1 having SEQ ID NO: 103, Ercc1 having SEQ ID NO: 104, MDC1 having SEQ ID NO: 105, Bard1 having SEQ ID NO: 106, Ligase 1 having SEQ ID NO: 107, Mre11 having SEQ ID NO: 108 and/or 53BP1 having SEQ ID NO: 109.

The recombinant mammalian cell may a primary stem cell, a hamster, e.g., CHO (Chinese hamster ovary), cell or a human, e.g., HEK293 cell.

The invention is also directed at a recombinant mammalian cell comprising:

a.) at least one TEP functional RNA and/or at least one recombinant nucleic acid sequence encoding a TEP protein or encoding TEP functional RNA ,
and
b) a recombinant nucleic acid molecule comprising:
(i) at least one transgene of interest, and
(ii) optionally, a MAR element.

As noted above, the TEP protein may be one or more of the following proteins of the protein secretion pathway: hSRP14, hSec61α1, hSec61β, hSec61γ, hSRP54, hSRP9, hSRPRα, hSRPβ, and hCANX.

The TEP protein may also correspond to one or more of the following amino acid sequences of proteins of the protein secretion pathway: hSRP14 having SEQ ID NO: 13, hSec61α1 having SEQ ID NO: 15, hSec61β having SEQ ID NO: 17, hSec61γ having SEQ ID NO: 19, hSRP54 having SEQ ID NO: 21, hSRP9 having SEQ ID NO: 23, hSRPRα having SEQ ID NO: 25, hSRPβ having SEQ ID NO: 27, and hCANX having SEQ ID NO: 29 and/or may correspond to amino acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

The TEP protein may be one or more of the following protein processing or metabolic proteins: hUCP4, hCMPSAT, rST6Gal1, hCOMSC, hT-Synthase, hP4HA1, hP4HB, hGILZ, hCyPB, hNRF2, hHK1, hPDI, hPIN1, hSEPW1, hCALR, hDDOST, hHSP40, hATP5A1, hSERCA2, hPDIA4, hHSC70 /HSPA8, hHYOU1, hCMP-SAS, hBeclin-1, hERdj3, CHO-AGE, hWip1, hRTP4, hREEP2, hDPM1 and hDRiP78.

The TEP protein may also correspond to one or more of the following amino acid sequences of protein processing or metabolic proteins: hUCP4 having SEQ ID NO: 31, hCMPSAT having SEQ ID NO: 33, rST6Gal1 having SEQ ID NO: 35, hCOMSC having SEQ ID NO: 37, hT-Synthase having SEQ ID NO: 39, hP4HA1 having SEQ ID NO: 41, hP4HB having SEQ ID NO: 43, hGILZ having SEQ ID NO: 45, hCyPB having SEQ ID NO: 47, hNRF2 having SEQ ID NO: 49, hHK1 having SEQ ID NO: 51, hPDI having

SEQ ID NO: 53, hPIN1 having SEQ ID NO: 55, hSEPW1 having SEQ ID NO: 57, hCALR having SEQ ID NO: 59, hDDOST having SEQ ID NO: 62, hHSP40 having SEQ ID NO: 64, hATP5A1 having SEQ ID NO: 66, hSERCA2 having SEQ ID NO: 68, hPDIA4 having SEQ ID NO: 70, hHSC70 /HSPA8 having SEQ ID NO: 72, hHYOU1 having SEQ ID NO: 74, hCMP-SAS having SEQ ID NO: 76, hBeclin-1 having SEQ ID NO: 78, hERdj3 having SEQ ID NO: 80, CHO-AGE having SEQ ID NO: 82, hWip1 having SEQ ID NO: 84, hRTP4 having SEQ ID NO: 86, hREEP2 having SEQ ID NO: 88, hDPM1 having SEQ ID NO: 90 and hDRiP78 having SEQ ID NO: 92 and/or may correspond to amino acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

The TEP protein may be a chaperone, in particular, a BiP protein, more in particular, a synthetic drosophila BIP protein derivative (DroBiP) having 80%, 90%, 95% or100% sequence identity with the SEQ ID NO.: 60.

The functional RNA may be a transiently transfected siRNA or an shRNA which is transcribed from said at least one isolated nucleic acid sequence, wherein the siRNA or the processed shRNA is a 20, 21, 22, 23, 24 or 25 base pairs long antisense RNA which is fully complimentary to 20, 21, 22, 23, 24 or 25 consecutive nucleotides of a mRNA of at least one target gene that is part of the NHEJ (non-homologous end-joining), HR (homologous recombination), MMEJ (Microhomolgoy mediated end joining) recombinant pathway or is a DNA repair protein such as MDC1 (mediator of DNA-damage checkpoint 1).

The at least one target gene that may be part of:

- the DNA repair and NHEJ is
53BP1 (Tumor suppressor p53-binding protein 1),
-the HR is
Rad51 (DNA repair protein RAD51), Rad51B (DNA repair protein RAD51 homolog 2), Rad51C (DNA repair protein RAD51 homolog 3),

Rad51D (DNA repair protein RAD51 homolog 4), Rad52 (DNA repair protein RAD52), Rad54 (DNA repair and recombination protein RAD54) ,
Xrcc2 (X-ray repair complementing defective repair in Chinese hamster cells 2),
Xrcc3 (X-ray repair complementing defective repair in Chinese hamster cells 3),
Brca1 (breast cancer 1, early onset),
Brca2 (breast cancer 2, early onset),
Bard1 (BRCA1 associated RING domain 1),
- the MMEJ is
Ercc1 (excision repair cross-complementing rodent repair deficiency, complementation group 1),
Mre11 (meiotic recombination 11)
Ligase1 (DNA ligase 1),
and/or
Is the DNA repair protein MDC1.

The target genes may be nucleic acids having at least 80%, 90%, 95%, 98% or 100% sequence identity with Rad51 having SEQ ID NO: 93, Rad51B having SEQ ID NO: 94, Rad51C having SEQ ID NO: 95, Rad51D having SEQ ID NO: 96, Xrcc2 having SEQ ID NO: 99, Xrcc3 having SEQ ID NO: 100, Rad52 having SEQ ID NO: 97, Rad54 having SEQ ID NO: 98 , Brca1 having SEQ ID NO: 101, Brca2 having SEQ ID NO: 102, Cyclin D1 having SEQ ID NO: 103, Ercc1 having SEQ ID NO: 104, MDC1 having SEQ ID NO: 105, Bard1 having SEQ ID NO: 106, Ligase 1 having SEQ ID NO: 107, Mre11 having SEQ ID NO: 108 and/or 53BP1 having SEQ ID NO: 109.

The at least one transgene may express a therapeutic protein such as an immunoglobulin, a hormone such as erythropoietin, or a growth factor and wherein, optionally, in the recombinant mammalian cell transgene integration and/or expression is increased relative to a cell not comprising said recombinant nucleic acid molecule(s).

The recombinant mammalian cell may comprise at least two TEP functional RNAs, wherein one or both of the TEP RNAs are transiently transfected siRNA, or are expressed by said isolated nucleic acid sequence(s) encoding a TEP functional RNA.

The recombinant mammalian cell may comprise a MAR element.

The invention is also directed at a method for transfecting mammalian cells, in particular hamster cells, comprising:
transfecting, optionally in a first transfection, said mammalian cells with

(i) at least one of said recombinant nucleic acids molecules of any one of claims 1 to 13 and/or

(ii) at least one isolated TEP functional RNA and at least one transgene which is, optionally, part of a recombinant nucleic acid molecule which is optionally transfected in a second, subsequent transfection, optionally together with an isolated nucleic acid or mRNA expressing a transposase which recognizes the 5' and the 3' ITR .

The recombinant mammalian cell may be transfected with more than one, including at least two, at least three or at least four of said recombinant nucleic acid molecules encoding one, two or three of the following: hSRP14, hSec61α1, hSec61β, hSec61γ, hSRP54, hSRP9, hSRPRα, hSRPβ, and hCANX.

I may also be transfected with hSRP14 having SEQ ID NO: 13, hSec61α1 having SEQ ID NO: 15, hSec61β having SEQ ID NO: 17, hSec61γ having SEQ ID NO: 19, hSRP54 having SEQ ID NO: 21, hSRP9 having SEQ ID NO: 23, hSRPRα having SEQ ID NO: 25, hSRPβ having SEQ ID NO: 27, and hCANX having SEQ ID NO: 29 and/or with amino acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

Any of the recombinant nucleic acid molecules may part of a vector, wherein the vectors may be co-transfected.

The co-transfectionof vectors encoding several TEP proteins, preferably a co-transfection of recombinant nucleic acid molecules encoding proteins SRP14, SRP9 and SRP54 may increase transgene integration and/or expression in said cell relative to a cell not subject to such co-transfection. Co-transfection of vectors comprising nucleic acid sequences have at least 80%, 90%, 95%, 98% or 100% SEQ ID NO: 12, SEQ ID NO: 22, SEQ ID NO: 20 is also within the scope of the present invention.

A number of said mammalian cells that stably express said TEP protein or TEP functional RNA may be obtained to obtain recombinant mammalian cells and wherein said number of recombinant mammalian cells may be independent from the presence of said MAR element. The mammalian cells may be transfected a second and optionally third time.
Preferably, at least 30%, 40% or 45% of said mammalian cells may become recombinant mammalian cells and express said transgene.

The mammalian cellmay be transfected with said at least one isolated TEP functional RNA and a vector comprising said at least one transgene and optionally a 3' ITR and 5' ITR flanking said at least one transgene, optionally together with an isolated nucleic acid or mRNA expressing a transposase with recognizes the 5' and the 3' ITR.

The transgene may be a therapeutic protein such as an immunoglobulin, hormone, cytokine or growth factor.

The recombinant nucleic acid molecule may comprise optionally a selection marker and wherein the at least one transgene is expressed
(a) without selection for said marker, or
(b) with selection for said marker, e.g. via a selection agent contained in a culture medium, and
(c) in absence of transposase, or
(d) in presence of transposase.

The TEP functional RNA may be encoded by a recombinant nucleic acid sequences encoding a shRNA or miRNA, or may comprise/consist of a siRNA that interferes with the expression of at least one HR gene, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2, Cyclin D1, Ercc1, MDC1, Bard1, Ligase 1, Mre11 and/or 53BP1.

The TEP functional RNA may also interfere with expression of genes having at least 80%, 90%, 95%, 98% or 100% sequence identity with Rad51 having SEQ ID NO: 93, Rad51B having SEQ ID NO: 94, Rad51C having SEQ ID NO: 95, Rad51D having SEQ ID NO: 96, Xrcc2 having SEQ ID NO: 99, Xrcc3 having SEQ ID NO: 100, Rad52 having SEQ ID NO: 97, Rad54 having SEQ ID NO: 98 , Brca1 having SEQ ID NO: 101, Brca2 having SEQ ID NO: 102, Cyclin D1 having SEQ ID NO: 103, Ercc1 having SEQ ID NO: 104, MDC1 having SEQ ID NO: 105, Bard1 having SEQ ID NO: 106, Ligase 1 having SEQ ID NO: 107, Mre11 having SEQ ID NO: 108 and/or 53BP1 having SEQ ID NO: 109.

The transgene integration and/or expression may be increased in such a cell relative to a cell not transfected with said isolated nucleic acid molecules and/or said at least one of said isolated TEP functional RNAs.

The invention is also directed at a kit comprising in one container at least one vector comprising the any one of the recombinant nucleic acid molecules according to claims 1 to 13 and, in a second optional container a vector encoding a compatible transposase and in a further container instruction of how to use the vector or vectors.

The kit mentioned above, wherein more than one vector is provided in one or more containers and wherein the TEP proteins are at least two of the following: a chaperone, SRP14, SRP9, SRP54, SR or a translocon.

The kit mentioned above, wherein the TEP functional RNA(s) within said vector(s) comprise(s)/consist(s) of nucleic acid sequences encoding a miRNA, siRNA or a shRNA

that interferes with the expression of at least one HR gene, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2-Cyclin D1, Ercc1, MDC1, Bard1, Ligase 1, Mre11 and/or 53BP1, and preferably in a further container siRNA(s) that interferes with the expression of at least one other HR gene, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2, Cyclin D1, Ercc1, MDC1, Bard1, Ligase 1, Mre11 and/or 53BP1.

The HR gene may correspond to nucleic acids having at least 80%, 90%, 95%, 98% or 100% sequence identity with Rad51 having SEQ ID NO: 93, Rad51B having SEQ ID NO: 94, Rad51C having SEQ ID NO: 95, Rad51D having SEQ ID NO: 96, Xrcc2 having SEQ ID NO: 99, Xrcc3 having SEQ ID NO: 100, Rad52 having SEQ ID NO: 97, Rad54 having SEQ ID NO: 98 , Brca1 having SEQ ID NO: 101, Brca2 having SEQ ID NO: 102, Cyclin D1 having SEQ ID NO: 103, Ercc1 having SEQ ID NO: 104, MDC1 having SEQ ID NO: 105, Bard1 having SEQ ID NO: 106, Ligase 1 having SEQ ID NO: 107, Mre 11 having SEQ ID NO: 108 and/or 53BP1 having SEQ ID NO: 109.

The invention is also directed towards the use of the recombinant nucleic acids disclosed herein and/or the recombinant mammalian cells disclosed herein, preferably for increasing transgene integration and/or expression.

The invention is also directed at an expression vector comprising:

(a) a transgene which is flanked, upstream by a promoter and downstream by a polyadenylation signal,

(b) a singular MAR element downstream of the polyadenylation signal, or

(c) a first MAR element upstream of the transgene of interest and a second MAR element downstream of said transgene integration site.

The singular or first and second MAR elements may be selected from MAR elements 1_68, 1_6, 1_6R2, 1_68R, 1_68R2, X_29R3 or X_29 or elements that have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NOs: 1, 2, 3, 6, 7, 8, 9 or 10.

The singular or first and second MAR(s) may be selected from rearranged MAR elements 1_6R2, 1_68R, 1_68R2 or X_29R3 or elements that have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NOs: 6, 7, 8, or 10, wherein, optionally, the MAR element(s) increase an expression of the transgene of interest by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70% relative to their non-rearranged counterparts.

The promoter may be an EF1 alpha promoter and the polyadenylation signal is a BGH polyA signal.

The vector may comprise promoter(s) and/or enhancer(s) or fusions thereof such as GAPDH, CGAPD, CSV40p, CMVp, CHO Ef1alpha, CHO Actb or CHO Hspa5.

The promoters may be GAPDH having SEQ ID NO: 111, CGAPDH having SEQ ID NO: 11, SV40p having SEQ ID NO: 114, CMVp having SEQ ID NO: 113, CHO EF1 alpha having SEQ ID NO:112, CHO Actb having SEQ ID NO: 115 and/or CHO Hspa5 having SEQ ID NO: 116 and nucleic acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

The promoter may be GAPDH promoter and comprises a CMV enhancer.

The first and/or second MAR, enhancer, promoter, transgene of interest and polyadenylation signal may be located between a 5' and 3' ITR.

In certain embodiment, the expression vector may comprise:

(a) a singular MAR element downstream of the polyadenylation signal, wherein said singular MAR element is preferably a MAR 1-68 or a MARX-29 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NOs. 1 or 3, in particular a rearranged MAR based on MAR 1-68 or a

MAR X-29, in particular an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID Nos : 6, 7 OR 10 (MARs 1_68R, 1_68R2 or X_29R3) or SEQ ID NO: 9, and is preferably a MAR X-29 element and/or has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 3, or

(b) a first MAR element upstream of the transgene of interest and a second MAR element downstream of said transgene of interest wherein the first MAR element, preferably comprises a 1_6 element and/or has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 2, in particular with rearranged MARs based on MAR 1-6, in particular elements that have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO: 8 (MAR 1_6R2) and the second MAR element that preferably comprises a MAR 1-68 element and/or has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO.1.

The expression vector may comprise a singular MAR element and the singular MAR element may be positioned downstream of the polyadenylation site and is a MAR 1-68 or a MAR X-29 and/or has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NOs. 1 or 3, in particular rearranged MARs based on MAR 1-68 or a MAR X-29, in particular elements at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID Nos : 6, 7 OR 10 (MARs 1_68R, 1_68R2 or X_29R3) or 9, and may be preferably a MAR X-29-derived element and/or has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 3.

The first MAR element may be upstream of the transgene of interest and a second MAR element downstream of said transgene of interest, wherein the first MAR element may comprise a MAR 1_6 element and/or may have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO 2 and the second MAR element may comprise a MAR 1_68 element and/or may have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 1.

The invention is also directed at a method for expressing a transgene comprising:

providing a recombinant mammalian cell comprising one of the vectors mentioned above comprising said transgene and expressing the transgene, wherein said MAR elements(s) may increase an expression of the transgene preferably by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70%. The vector may comprise a singular MAR X_29 element and/or a nucleic acid that may have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 3 and wherein, after more than 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 weeks of culture, the MAR element may increase an expression of the transgene of interest by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70%.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1. Transposon Vector Construction

To test whether adding a MAR element to the PB (PiggyBack) transposon may affect transposition efficiency and transgene expression, and to assess whether the location of the MAR in the construct had any influence on these effects, a series of transposon donor constructs containing the GFP and puromycin resistance (Puro) gene were designed, in which the MAR 1_68 or a control neutral spacer DNA sequence were inserted at different positions in the plasmid. The parental Puro-GFP transposon plasmid without an insert was used as a control of transposition, to distinguish the impact of increased transposon size relative to effect of the MAR or spacer sequence addition.

Fig. 2 Transposon Vectors: Transposition Efficiency

Transposition efficiency of the various transposon constructs was measured by assessing the (A) percentage of GFP-expressing cells after transfection and three weeks of cultivation without antibiotic selection and (B) by counting puromycin-resistant colonies.

Fig. 3 Transposon Vectors: Expression Level

Analysis of the expression level allowed by different Transposon vectors tranfected with (+PB) or without (-PB) transposase expression plasmid, by probing the GFP fluorescence levels of the CHO cells after 3 weeks of culture without (A) or with (B) secretion for puromycin resistance following the transfection, taking into account the fluorescence of GFP-positive cells only.

Fig. 4 Effect of the MAR and transposase on transgene genomic integration

The number of integrated GFP transgene copies was determined using qPCR, and values were normalized relative to the cellular B2M gene, using genomic DNA isolated from unselected CHO cells (A), or puromycin-resistant cells (B) generated as described in the legends to Figs. 1-3. Values represent the means ± SEM (n=3). *P<0,05.

Fig. 5 Transgene expression per transgene

Assessment of the intrinsic expression potential of the vectors, independently from their propensity to integrate in the genome, without (A) and with (B) puromycin selection.

Fig. 6 Effect of the expression of secretion proteins from transposable and plasmid vectors on recombinant protein (transgene) expression. Transposable or regular plasmid vectors were constructed to express secretion proteins SRP9, SRP14, SRP54, the SRP receptor alpha and beta subunits (SR), or the Translocon. Transposable vectors were co-transfected with the PiggyBack transposase vector **(right panel)**, whereas the non-transposable plasmid vectors were transfected alone **(left panel)**, in a cell clone expressing the Infliximab antibody as described herein. After three weeks of culture with selection (left panel) or without selection (right panel), the levels of secreted infliximab antibody were assayed from cell culture supernatants. As can bee seen, the specific productivity, that is the relative expression of a cell containing a sequence encoding a transgene expression processing (TEP) protein or TEP functional RNA was increased when using the transposon vector, from between 0.25 - 1.5 to between 1- 2.5, respectively, relative to the parental cell without a TEP.

Fig. 7: Recombinant protein expression from electroporated CHO-M cell suspensions

(A) CHO-M cells that were electroporated once or twice with the MAR X-29-bearing GFP-expression transposon vector in the presence (+PB) or absence (-PB) of the piggyBac transposase. The percentage of stable GFP-expressing cells after 3 weeks of culture performed in the absence of selection is shown.

(B) Mean of the GFP fluorescence of the GFP-positive cells.

(C) cDNAs encoding immunoglobulin light and heavy chains of the Bevacizumab (Beva), Adalimumab (Adal) and Rituximab (Ritu) antibody were introduced in MAR X29-containing transposon plasmids instead of GFP. The light and heavy chain transposon

constructs were electroporated three times at 12 days intervals with the piggyBac transposase expression vector in CHO-M cells. The levels of immunoglobulin secreted in the culture supernatants of polyclonal cell pools grown without selection is shown (open bars). Alternatively, the unselected polyclonal cell populations were sorted by panning cells displaying immunoglobulins at their surface using magnetic micro-beads: The levels of secreted immunoglobulins for the unsorted populations are shown (closed bars).

(D) Immunoglobulin-expressing colonies were sorted from transfected cell populations using a colony-picking device, and two clones expressing each of the three immunoglobulins were grown in fed-batch cultures in spin-tube bioreactors. The levels of secreted immunoglobulins are shown and were determined as for panel (C).

Fig. 8 Heterologous expression of SRP14 improves Trastuzumab secretion and restores Infliximab secretion

CHO-K1 HP and LP clones expressing the Trastuzumab (A) or Infliximab (B) immunoglobulins at the highest obtained levels, stably re-transfected with the SRP14 expression vector and monoclonal populations were isolated. The derived subclones, labeled A to E, were evaluated for cell growth and production in batch culture conditions. Cell density (cells/ml) and IgG titer (μg/ml) were plotted for each sampling day through the 7 days of culture. (C) Specific productivity distribution of the TrastuzuMab (HP) and InflixiMab (LP) subclones after transfection with the SRP14 expression vector (lanes S) as compared to that of the parental HP and LP clones (-). (D) The relative levels of SRP14 mRNA was determined for the 5 individual SRP14-LP A-E subclones and the parental control LP clone, and they were plotted relative to the specific IgG productivity from 4 culture runs. mRNA and specific productivity mean and standard deviation values are expressed as the fold increase over those of the LP control clone.

Fig. 9 Heterologous expression of SRP14 mediates high yield of the hard-to-express immunoglobulin in a production process

The SRP14 vector-transfected TrastuzuMab HP subclone B (A) and InflixiMab LP subclone E (B), as analysed in Fig. 8, were cultivated in 125 ml ventilated shake flask vessel with a working volume of 25 ml in fed-batch cultures, and the viable cell density and IgG titer were determined during an 11-days time course.

Fig. 10 SRP14 expression abolishes light chain aggregation by CHO cell clones

(A) The supernatants and pellets of Tx100-permeabilized cells collected by centrifugation were analyzed by SDS-PAGE, as depicted by the Tx-100 soluble and Tx-100 insoluble labeled panels, respectively, for the LP-derived SRP14-LP subclone E and the HP-derived SRP14-HP subclone B (lanes S), or for CHO subclones expressing a control GFP protein (lanes G). Arrowheads show the misprocessed free LC and aggregated (Aggr.) LC. (B) Chase analysis of the various LC, HC and IgG assembly intermediates species produced by SRP14-LP clone E and LP-control clone E was performed and results are shown.

Fig. 11 Effect of combined expression of SRP, SR and translocon subunits on immunoglobulin secretion

(A) An Infliximab LP clone E was re-transfected with various combinations of SRP, SR and translocon transposable expression vectors. The specific productivity of the resulting cell pools was then evaluated in batch cultivation and represented as a % of the LP-control cells pcd values. Box-plots represent the median, upper and lower quartiles of the normalized specific productivities determined at day 3 of independent culture runs. (B) The SRP14-expressing infliximab producing cell subclone E was re-transfected with various SR and translocon transposable expression vector combinations. The specific productivity of cell pools is represented as for panel A.

Fig. 12 Model for the rescue of Infliximab secretion from SRP14-expressing clones

Model of the IgG folding and secretion by low producer clones before (A) and after SRP/Translocon subunits overexpression (B). The data indicate that neosynthetized LC produced by low producer clones exhibit improper processing and folding state. Signal peptide misprocessing of the Infliximab LC may lead to the saturation of the ER co-translational translocation machinery (panel A, number 1). Its aggregation in the ER within IgG assembly-incompetent aggregated LC forms (panel A, number 2) induce ER stress and trigger the formation of autophagosome-like structure (panel A, number 3). Overexpression of the SRP14 and others SRP/translocon components proteins fully rescued the processing and secretion of the InflixiMab IgG (panel B). SRP14 elongation arrest activity possibly delays LC ER translocation during translation of its mRNA (panel B, number 1). This would favor in turn the correct processing of the LC and its proper interaction with ER folding chaperones (panel B, number 2). The maintenance of the neosynthetized LC in an IgG assembly-competent state thus restores high yield secretion of fully-assembled antibodies (panel B, number 3).

Fig. 13 Effect of si-RNA knock-down of HR and NHEJ on expression

Fold differences in the percentage of GFP-positive cells (with respect to cells transfected with a GFP control plasmid shown here as 1.0) representing the frequencies of recombination events in untreated cells (mock), cells treated with negative siRNA (siNeg), siRNAs against NHEJ factors (siKu70+80+DNA-PKcs) and anti-HR siRNA (siRad51). The GFP lanes show a positive control of GFP expressing cells. The HR undigest. and NHEJ undigest. - labeled lanes show negative control cells, i.e. cells transfected with circular HR and NHEJ report plasmids. The HR I-SceI and NHEJ I-SceI-labelled lanes indicate cells transfected with SceI-cleaved reporter plasmids that restore GFP expression upon DNA cleavage repair by homologous recombination or non-homologous end joining, respectively. The figure shows the efficacy of the siRNA to inhibit HR or NHEJ, as indicated by the percent of GFP-positive cells, which was normalized to the percent of dsRed-positive cells and expressed as the fold change over the percentage of the GFP control cells, which was set to 1. Mean of 3 experiments, error bars show standard error of the mean. Statistical significance determined by unpaired Student's t-test; significance level $p<0.05$ (*) and $p<0.01$ (**).

Fig. 14 Effect of MARs in siRNA knock-down of NHEJ

The fold increase in GFP expression and integration CHO cells treated with siRNAs against NHEJ factors and retransfected with a GFP or MAR-GFP plasmids is shown. The average GFP fluorescence, copy number and fluorescence per GFP copy is shown as a fold increase over the result obtained from untreated cells (marked as 'mock') transfected with the GFP plasmid. A) Flow cytometry results, B) analysis of GFP copy number in the genome by qPCR, C) average fluorescence of each integrated GFP gene (calculated for each experiment as a ratio between expression and copy number). Mean of 3 or more experiments; statistical significance determined by unpaired Student's t-test. Asterisks indicate significant differences between the siRNA-treated sample and corresponding untreated control; significance levels: $p<0.05$ (*), $p<0.01$ (**); error bars show standard error of the mean.

Fig. 15 Effect of MARs in siRNA knock-down of HR

The fold increase in GFP expression and integration CHO cells treated with siRNAs against HR factors and retransfected with a GFP or MAR-GFP plasmids. The average GFP fluorescence, copy number and fluorescence per GFP copy is shown as a fold increase over the result obtained from untreated cells (marked as 'mock') transfected with the GFP plasmid. A) Flow

cytometry results, B) analysis of GFP copy number in the genome by qPCR, C) average fluorescence of each integrated GFP gene (calculated for each experiment as a ratio between expression and copy number). Mean of 3 or more experiments; statistical significance determined by unpaired Student's t-test. Asterisks indicate significant differences between the siRNA-treated sample and corresponding untreated control; significance levels: $p<0.05$ (*), $p<0.01$ (**); error bars show standard error of the mean.

Fig. 16 Effect of MARs in siRNA knock-down of MMEJ

GFP expression and integration CHO cells treated with siRNAs against MMEJ factors (and some HR factors) and retransfected with a GFP **(A)** or MAR-GFP plasmids **(B)** is shown. The average GFP fluorescence, copy number and fluorescence per GFP copy is shown as a fold increase over the result obtained from untreated cells (marked as 'mock') transfected with the GFP plasmid. The figures show the flow cytometry results. Shown is the mean of the number of experiments indicated at the bottom. Cells transfected with siMDC1, expressed GFP even without MAR, at a 11.8 higher rate as cells not transfected with siMDC. Particularly good results could also be achieved with certain plasmids that did contain MAR, namely siBard1 and siLigI.

Fig. 17 Effect of si-RNA-mediated knock-down of a HR protein

The figure shows that higher GFP and immunoglobulin expression can be achieved from CHO-M cells stably expressing a Rad51-directed shRNA. CHO-M cells were transfected with a PiggyBac-derived transposable Rad51 shRNA expression vector, and the polyclonal cell pool as well as cell clones derived thereof were retransfected with a GFP expression plasmid along with the parental CHO-M cells. The GFP fluorescence of the parental CHO-M, of the Rad51-shRNA expressing cell pool and of the derived clones was assessed 10 days after selection for stable expression of the GFP and puromycin resistance genes. The fluorescence profiles of two of the most fluorescent clones are shown next to those of the cell pool and parental cells (A), as well as the percentage of cells in the M1, M2 and M3 sectors 10 days after selection for puromycin resistance (B), as depicted by the horizontal bars labeled 1, 2 and 3 in panel A. The proportion of highly expressing M3 cells was followed during 68 days of further culture without selection to show that higher and more stable expression can be obtained from the shRNA-expressing cell clones when compared to the parental CHO_M cells (C). Alternatively, an expression plasmids encoding the light and heavy chains of the Infliximab antibody were transfected into representative clones, and the specific productivity of secreted immunoglobulin was assessed after selection during three weeks of further culture without antibiotic.

Fig. 18 Effect of various human recombinant upstream MARs on the percentile of high and very high producer cells (% M3/M2), as assessed for GFP fluorescence by FACS analysis in a two MAR construct. (A) The MAR elements were rearranged derivatives of MAR X-29 (X_29R2 (SEQ ID NO: 9), X_29R3 (SEQ ID NO: 10), MAR 1-42 (1_42R2Bis, 1_42R3), MAR 1-6 (1_6R2 (SEQ ID NO: 8), 1_6R3) or MAR 1-68 (1_68R2 (SEQ ID NO: 7), as indicated in the names of the constructs. (B) Typical FACS profiles obtained for the best upstream MAR elements (MAR 1_68R (SEQ ID NO: 6)).

Fig. 19 Stability of expression in a two MAR vector

Polyclonal populations constructed from vectors containing the 1_68R2, 1_6R2 and X_29R3 MAR derivatives was tested over a period of 5 weeks of culture without selection and GFP fluorescence was assessed weekly over this period. The percentile of the M3 subpopulation were assessed: 1_6R2 element as the upstream MAR and the unrearranged MAR 1-68 as downstream MAR were the best tested combination of vector with two MARs. M1 and M2 subpopulations are also shown.

Fig. 20 Expression vectors containing a single genetic element

MAR 1_68 and X_29 were tested and used in combination with the LmnB2 replicator. The MARs were positioned downstream the transgene expression cassette and were assessed in transgene transfection assay over a period of two months. The polyclonal population of stably transfected cells was selected for antibiotic resistance during two weeks and tested for GFP fluorescence by fluorescence-activated cell sorter (FACS) analysis during seven weeks. The proportion of high producer M3 cells is shown in (A), while typical FACS profiles are shown in (B).

Fig. 21 Expression vectors containing a single genetic element: X-29

Stability assay of the X_29 vector: The expression vector containing a single X_29 downstream the expression cassette is shown to be stable and to give a very high percentile of M2 and M3 subpopulations even after 14 weeks of culture (27 passages).

Fig. 22 Comparative analysis of stably transfected CHO populations after 24 weeks of antibiotic selection

A vector with a single X_29 MAR downstream the expression cassette

(Puro_CGAPD_GFP_gastrin_X29) increases the occurrence of high GFP expressing cells and also the stability of the expression over time compare to the vector with two MARS with 1_6R2 as upstream MAR and 1_68 as downstream MAR (Puro_1_6R2_CGAPD_GFP_gastrin_1_68).

DETAILED DESCRIPTION OF VARIOUS AND PREFERRED EMBODIMENTS

A **transgene** as used in the context of the present invention is an isolated deoxyribonucleotide (DNA) sequence coding for a given mature protein (also referred to herein as a **DNA encoding a protein**), for a precursor protein or for a **functional RNA** that does not encode a protein (non-coding RNA). A transgene is isolated and introduced into a cell to produce the transgene product. Some preferred transgenes according to the present invention are transgenes encoding immunoglobulins (Igs) and Fc-fusion proteins and other proteins, in particular proteins with therapeutical activity ("biotherapeutics"). For instance, certain imunoglobulins such as Infliximab (Remicade) or other secreted proteines such as coagulation factor VIII, are notably difficult to express, because of mostly uncharacterized cellular bottlenecks. With the help of the recombinant nucleic acid molecules, vectors and methods of the present invention these bottlenecks may be identified and/or opened. This generally increases the amount of therapeutic proteins that can be produced and/or their quality, such as e.g. their processing and the homogeneity of post-translational modifications such as glycosylation.

As used herein, the term transgene shall, in the context of a DNA encoding a protein, not include untranscribed flanking regions such as RNA transcription initiation signals, polyadenylation addition sites, promoters or enhancers. Other preferred transgenes include DNA sequences encoding functional RNAs. Thus, the term transgene is used in the present context when referring to a DNA sequence that is introduced into a cell such as an eukaryotic host cell via transfection (which includes in the context of the present invention also transduction, i.e., the introduction via viral vectors) and which encodes the product of interest also referred to herein as the "**transgene expression product**", e.g., "**heterologous proteins**". The transgene might be functionally attached to a signal peptide coding sequence, which encodes a signal peptide which in turn mediates and/or facilitates translocation and/or secretion across the endoplasmic reticulum and/or cytoplasmic membrane and is removed prior or during secretion.

Small interfering RNAs (**siRNA**) are double stranded RNA molecules, generally 20-25 base pairs long which play a role in RNA interference (RNAi) by *interfering with the expression of specific genes with complementary nucleotide sequence*. A siRNA can be directly introduced into the cells or can be expressed in the cell via a vector. An isolated TEP siRNA as referred to

herein is such a 20-25 base pair long siRNA that is usually introduced directly into the cell, i.e., without being expressed via a nucleic acid that has been introduced into the cell.

A small/short hairpin RNA (shRNA) is a sequence of RNA that makes a tight hairpin turn that can be used to silence target gene expression via RNAi. Expression of shRNA in cells is typically accomplished by delivery of plasmids or viral vectors such as retroviral vectors. To create shRNAs, a siRNA sequence is usually modified to introduce a short loop between the two strands of the siRNA. A nucleic acid encoding the shRNA is then delivered via a vector into the cell and are transcribed into short hairpin RNA (shRNA), which can be processed into a functional siRNA by Dicerin its usual fashion.

An si/shRNA is capable of sequence-specifically reducing expression of a target gene. The shRNA may hybridize to a region of an mRNA transcript encoding the product of the target gene, thereby inhibiting target gene expression via RNA interference. Bi-functional shRNAs have more than one target, e.g., the coding region as well as certain untranslated regions of an mRNA. Integration into the cell genome facilitates long-lasting or constitutive gene silencing that may be passed on to progeny cells.

A microRNA (miRNA) is a small RNA molecule, e.g., 20 to 24, in particular 22 nucleotides long, which functions in transcriptional and post-transcriptional regulation of gene expression via pairing with complementary sequences within mRNAs. Gene silencing may occur either via transgene transcription inhibition, mRNA degradation or preventing mRNA from being translated. miRNAs can be expressed by delivery of plasmids or viral vectors such as retroviral vectors. Alternatively, RNA molecules inhibiting or mimicking miRNA can by synthesized and transfected directly in cells.

A "**Sequence encoding a transgene expression processing (TEP) protein or TEP functional RNA**" **allows the expression or the increased expression of the given TEP protein following its transfer into a cell, whereas the sequence encoding a** non-coding functional RNAs inhibit the expression of cellular proteins, respectively. The TEP proteins can be identical or similar to cellular proteins, or they can be proteins from a distinct cell or species. The cellular proteins whose expression is, e.g., inhibited by functional RNAs are constituent proteins of the cell into which functional RNAs are introduced. The TEP protein may also supplement the expression of another cellular protein and as a result, preferably, enhance the expression of a transgene. The proteins may be involved in recombination; in mRNA translational processes; in ER translocation, secretion, processing or folding of polypeptides, in ER-Golgy-plasma membrane transport, glycosylation and/or another post-translational modification. Functional RNAs include, e.g., siRNAs, shRNAs, microRNAs, lariat-form spliced

RNA, short-temporary sense RNA (stRNA), antisense RNA (aRNA), ribozyme RNA and other RNAs, in particular those that can knock-down target gene expression. In a particular preferred embodiment, these proteins are involved in the "**The Protein secretion pathway**" **or in "The Recombination pathways",** but also include certain protein processing or metabolic proteins as described below.

TEP functional RNAs may not only be expressed from a nucleic acid sequence as described above, but may be directly introduced into the cell. This, in particular is true for isolated TEP siRNAs.

The term an "isolated nucleic acid molecule" is in the context of the present invention is equivalent to a "recombinant nucleic acid molecule", i.e., a nucleic acid molecule that, does not exist, in this form in nature, but has been constructed starting from parts that do exist in nature.

A nucleic acid sequence, such as a DNA or RNA, is complimentary to another DNA or RNA, if the nucleotides of, e.g., two single stranded DNA stands or two single stranded RNA strands can form stable hydrogen bonds, such as a hydrogen bond between guanine (G) with cytosine (C). In the cell, complementary base pairing allows, e.g., cells to copy information from one generation to another. In RNA interference (RNAi) complementary base pairing allows, the silencing or complete knock-out of certain target genes. Essentially, siRNA, shRNA or miRNA sequence specifically reduce or knock-out expression of a target gene by having a single RNA strand (e.g. the anti-sense strand in siRNA) align with RNA, in particulary the mRNA of the host cell. The degree of complementarity between two nucleic acid strands may vary, from complete complementarity (each nucleotide is across from its opposite) to partial complementary (50% , 60%, 70%, 80%, 90% or 95%). The degree of complementarity determines the stability of the complex and thus how successfully a gene can be, e.g., knocked-out. Thus, complete or at least 95% complementarity are preferred.

The activity of siRNAs in RNAi is largely dependent on its binding ability to the RNA-induced silencing complex (RISC). Binding of the duplex siRNA to RISC is followed by unwinding and cleavage of the sense strand with endonucleases. The remaining anti-sense strand-RISC complex can then bind to target mRNAs for initiating transcriptional silencing.

Within the context of the present invention **transgenes**, as defined above, express generally proteins whose production in larger quantities is desired, e.g. for pharmaceutical use, while sequences encoding TEP proteins/functional RNAs, or the functional RNAs themselves, are

designed to help the expression of such transgenes either directly or indirectly. An "exemplary list of TEP proteins expressed using transposon vectors" is listed as **TABLE A**. As the person skilled in the art will appreciate, the huge majority of these proteins have been disclosed in the art and Table A discloses both the NCBI reference sequence numbers for the respecitve proteins as well as the nucleic acid sequence encoding the same. The last column provides sequence identifiers for certain of those sequences. The person skilled in the art will apprecitate that variants of the proteins as well a sequences with more then 80%, 90%, 95% or 98% sequence identity are part of the present invention.

An "exemplary list of shRNA expressed using, e.g., specific piggybac transposon vectors" is listed as **TABLE B**. As the person skilled in the art will appreciate, such shRNAs can be readily constructed when a target gene has been selected. For example any one of the known genes of the recombination pathway is a ready target gene. However, other genes, such as genes for the proteins set forth in **Table A** may be ready targets for siRNAs generated from those shRNAs. **TABLE C** is a list of examples of siRNAs (sense strand) and examples of shRNAs created from corresponding siRNAs. The antisense strand of the siRNA is ultimately used to block and/or provoke the degradation of a cellular mRNA. This generally leads to reduced levels of the protein encoded by the mRNA.

Identity means the degree of sequence relatedness between two nucleotide sequences as determined by the identity of the match between two strings of such sequences, such as the full and complete sequence. Identity can be readily calculated. While there exists a number of methods to measure identity between two nucleotide sequences, the term "identity" is well known to skilled artisans (Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Methods commonly employed to determine identity between two sequences include, but are not limited to those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipman, D., SIAM J Applied Math. 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the two sequences tested. Such methods are codified in computer programs. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG (Genetics

Computer Group, Madison Wis.) program package (Devereux, J., et al., Nucleic Acids Research 12(1). 387 (1984)), BLASTP, BLASTN, FASTA (Altschul et al. (1990); Altschul et al. (1997)). The well-known Smith Waterman algorithm may also be used to determine identity.

As an illustration, by a nucleic acid comprising a nucleotide sequence having at least, for example, 95% "identity" with a reference nucleotide sequence means that the nucleotide sequence of the nucleic acid is identical to the reference sequence except that the nucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a nucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Sequence identities of more about 60%, about 70%, about 75%, about 85% or about 90% for any sequence disclosed herein (e.g., by SEQ IDs. And/or by accession numbers) are also within the scope of the present invention.

A nucleic acid sequence having **substantial identity** to another nucleic acid sequence refers to a sequence having point mutations, deletions or additions in its sequence that have no or marginal influence on the respective method described and is often reflected by one, two, three or four mutations in 100 bps.

The invention is directed to both polynucleotide and polypeptide variants. A "**variant**" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide disclosed, but retaining essential properties thereof. Generally, variants are overall closely similar and in many regions, identical to the polynucleotide or polypeptide of the present invention.
The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids disclosed

herein are substituted, deleted, or added in any combination are also preferred.
The invention also encompasses allelic variants of said polynucleotides. An allelic variant denotes any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene. Variants of any of nucleic acid molecules disclosed herein are part of the present invention.

A **promoter** sequence or just promoter is a nucleic acid sequence which is recognized by a host cell for expression of a specific nucleic acid sequence. The promoter sequence contains transcriptional control sequences which regulate the expression of the polynucleotide. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell. Promoters according to the present invention include inducible and non-inducible promoters. A nucleic acid sequence is **under control** of a promoter is the promoter exercises its function on said nucleic acid.

CGAPDH (also referred to herein as C_GAPDH) is an enhancer-promoter fusion, which comprises the human GAPDH promoter and the human CMV immediate early gene enhancer. In one embodiment, to produce it, the human GAPDH promoter and its 5'UTR were PCR amplified from human HEK293 cell genomic DNA. The product was placed downstream of the human CMV immediate early gene enhancer. See SEQ ID NO: 11 for a representative sequence. Sequences having at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO. 11 are also within the scope of the present invention. Other desirable promoter(s) and/or enhancer(s) or fusions thereof are, but not limited to, the CMV IE enhancer, the human GAPDH promoter, the human Ef1 alpha promoter, the CMV promoter, the SV40 promoter, the CHO Actb promoter or the CHO Hspa5 promoter. These elements are well known in the art and sample sequences are listend under SEQ ID NOs: 110 to 116. As the person skilled in the art will understand, variants thereof are also part of the present inventions and well as element that have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with any one of SEQ ID NOs: 110 to 116.

A "**transposon**" is a mobile genetic element that efficiently transposes between vectors and chromosomes via a "cut and paste" or "copy and paste" mechanism. During transposition, the transposase (e.g., the PB transposase in the PiggyBac transposon system) recognizes transposon-specific inverted terminal repeat sequences (**ITRs**) located on both ends of the transposon (there is a 5'- and a 3' ITR to any transposon system) and moves the contents from the original sites and integrates them into chromosomal sites, such as TTAA chromosomal sites. The powerful activity of the PiggyBac transposon system enables genes of interest between the two ITRs to be easily mobilized into target genomes. The PiggyBac transposon system is described, e.g., in 2010/0154070, which is incorporated herein by reference in its entirety.

MAR elements (MAR constructs, MAR sequences, S/MARs or just MARs) belong to a wider group of epigenetic regulator elements which also include boundary or insulator elements such as cHS4, locus control regions (LCRs), stabilizing and antirepressor (STAR) elements, ubiquitously acting chromatin opening (UCOE) elements or histone modifiers such as histone deacetylase (HDAC).

MAR elements may be defined based on the identified MAR they are primarily based on: A MAR S4 **construct** is, accordingly, a MAR elements that whose majority of nucleotide (50% plus, preferably 60%, 70% or 80%) are based on MAR S4. Several simple sequence motifs such as high in A and T content have often been found within MARs. Other motifs commonly found are the A-box, the T-box, DNA unwinding motifs, SATB1 binding sites (H-box, A/T/C25) and consensus topoisomerase II sites for vertebrates or Drosophila.

MARs are generally characterized as sequences in the DNA of eukaryotic chromosomes where the nuclear matrix attaches. The properties of MAR are only in part defined by their primary structure. For example, a typical primary structure found in MAR elements such as AT rich regions are known to result in tertiary structures, namely in certain curvatures that define the function of the MAR. Thus, MARs are often defined not only by their primary structure, but also by their secondary, tertiary structure, e.g. their degree of curvature and/or physical properties such as melting temperature.

An AT/TA-dinucleotide rich bent DNA region (hereinafter referred to as "**AT-rich region**") as commonly found in MAR elements is a bent DNA region comprising a high number of A and Ts, in particular in form of the dinucleotides AT and TA. In a preferred embodiment, it contains at least 10% of dinucleotide TA, and/or at least 12% of **dinucleotide AT** on a stretch of 100 contiguous base pairs, preferably at least 33% of **dinucleotide TA,** and/or at least 33% of

dinucleotide AT on a stretch of 100 contiguous base pairs (or on a respective shorter stretch when the AT-rich region is of shorter length), while having a bent secondary structure. However, the "AT-rich regions" may be as short as about 30 nucleotides or less, but is preferably about 50 nucleotides, about 75 nucleotides, about 100 nucleotides, about 150, about 200, about 250, about 300, about 350 or about 400 nucleotides long or longer.

Some binding sites are also often have relatively high A and T content such as the SATB1 binding sites (H-box, A/T/C25) and consensus Topoisomerase II sites for vertebrates (RNYNNCNNGYNGKTNYNY) or Drosophila (GTNWAYATTNATNNR). However, a **binding site region** (module), in particular a TFBS region, which comprises a cluster of binding sites, can be readily distinguished from AT and TA dinucleotides rich regions ("AT-rich regions") from MAR elements high in A and T content by a comparison of the bending pattern of the regions. For example, for human MAR 1_68, the latter might have an average degree of curvature exceeding about 3.8 or about 4.0, while a TFBS region might have an average degree of curvature below about 3.5 or about 3.3. Regions of an identified MAR can also be ascertained by alternative means, such as, but not limited to, relative melting temperatures, as described elsewhere herein. However, such values are specie specific and thus may vary from specie to specie, and may, e.g., be lower. Thus, the respective AT and TA dinucleotides rich regions may have lower degrees of curvature such as from about 3.2 to about 3.4 or from about 3.4 to about 3.6 or from about 3.6 to about 3.8, and the TFBS regions may have proportionally lower degrees of curvatures, such a below about 2.7, below about 2.9, below about 3.1, below about 3.3. In SMAR Scan II, respectively lower window sizes will be selected by the skilled artisan.

A MAR element, a MAR construct, a MAR sequence, a S/MAR or just a MAR according to the present invention is a nucleotide sequence sharing one or more (such as two, three or four) characteristics such as the ones described above with a **naturally occurring "SAR" or "MAR"** Preferably such a **MAR element,** a MAR construct, a MAR sequence, a S/MAR or just a MAR has at least one property that facilitates protein expression of any gene influenced by said MAR. A **MAR element** has generally also the feature of being an isolated and/or purified nucleic acid preferably displaying MAR activity, in particular, displaying transcription modulation, preferably enhancement activity, but also displaying, e.g., expression stabilization activity and/or other activities.

The terms **MAR element,** MAR construct, a MAR sequence, a S/MAR or just a MAR also includes, in certain embodiments, **enhanced MAR constructs** that have properties that constitute an enhancement over an natural occurring and/or identified MAR on which a **MAR construct** according to the present invention may be based. Such properties include, but are not limited to, reduced length relative to the full length natural occurring and/or identified MAR, gene expression/transcription enhancement, enhancement of stability of expression, tissue specificity, inducibility or a combination thereof. Accordingly, a **MAR element** that is enhanced may, e.g., comprise less than about 90%, preferably less than about 80%, even more preferably less than about 70%, less than about 60%, or less than about 50% of the number of nucleotides of an identified MAR sequence. A **MAR element** may enhance gene expression and/or transcription of a transgene upon transformation of an appropriate cell with said construct.

A MAR element is preferably inserted upstream of a promoter region to which a gene of interest is or can be operably linked. However, in certain embodiments, it is advantageous that a MAR element is located upstream as well as downstream or just downstream of a gene/nucleotide acid sequence of interest. Other multiple MAR arrangements both in cis and/or in trans are also within the scope of the present invention.

Synthetic, when used in the context of a MAR element refers to a MAR whose design involved more than simple reshuffling, duplication and/or deletion of sequences/regions or partial regions, of identified MARs or MARs based thereon. In particular, synthetic MARs/MAR elements generally comprise one or more, preferably one, region of an identified MAR, which, however, might in certain embodiment be synthesized or modified, as well as specifically designed, well characterized elements, such as a single or a **series of TFBSs**, which are, in a preferred embodiment, produced synthetically. These designer elements are in many embodiments relatively short, in particular, they are generally not more than about 300 bps long, preferably not more than about 100, about 50, about 40, about 30, about 20 or about 10 bps long. These elements may, in certain embodiments, be multimerized. Such synthetic MAR elements are also part of the present invention and it is to be understood that generally the present description can be understood that anything that is said to apply to a "MAR element" equally applies to a synthetic MAR element.
Functional fragments of nucleotide sequences of identified MAR elements are also included as long as they maintain functions of a MAR element as described above.

Some preferred identified MAR elements include, but are not limited to, MAR 1_68, MAR X_29, MAR 1_6, MAR S4, MAR S46 including all their permutations as disclosed in WO2005040377 and US patent publication 20070178469, which are specifically incorporated by reference into the present application for the disclosure of the sequences of these and other MAR elements. The chicken lysozyme MAR is also a preferred embodiment (see, US Patent No. 7,129,062, which is also specifically incorporated herein for its disclosure of MAR elements).

If a vector is said to comprise **a singular** MAR this means that in this vector there is one MAR and there are no other MARs within the vector either of the same or a different type or structure.

In certain embodiments of the invention, there are multiple MARs, which maybe of the the same or a different type or structure and which may all be located downstream of a gene of interest. This is called a **singular MAR cluster**.

If something, such as a number of cells stably expressing a polypeptide, is said to be "**independent**" from the presence of, e.g., a sequence, then the sequence does not influence (e.g., the number of cells stably expressing a polypeptide) to any statistically significant extent.

A transgene or sequence encoding a transgene expression processing protein or functional RNA of the present invention is often part of a vector.

A **vector** according to the present invention is a nucleic acid molecule capable of transporting another nucleic acid, such as a transgene that is to be expressed by this vector, to which it has been linked, generally into which it has been integrated. For example, a plasmid is a type of vector, a retrovirus or lentivirus is another type of vector. In a preferred embodiment of the invention, the vector is linearized prior to transfection. An expression vector comprises regulatory elements or is under the control of such regulatory elements that are designed to further the transcription and/or expression of a nucleic acid sequence carried by the expression vector. Regulatory elements comprise enhancers and/or promoters, but also a variety of other elements described herein (see also "Vector Design").

The **vector sequence of a vector** is the DNA or RNA sequence of the vector excluding any "other" nucleic acids such as transgenes as well as genetic elements such as MAR elements.

An eukaryotic, including a mammalian cell, such as a recombinant mammalian cell/ eukaryotic host cell, according to the present invention is capable of being maintained under cell culture conditions. Non-limiting examples of this type of cell are non-primate eukaryotic host cells such as Chinese hamster ovary (CHOs) cells and baby hamster kidney cells (BHK, ATCC CCL 10). Primate eukaryotic host cells include, e.g., human cervical carcinoma cells (HELA, ATCC CCL 2) and monkey kidney CV1 line transformed with SV40 (COS-7, ATCC CRL-1587). A recombinant eukaryotic host cell or recominant mammalian cell signifies a cell that has been modified, e.g., by transfection with, e.g., a transgenic sequence and/or by mutation. The eukaryotic host cells or recominant mammalian cells are able to perform post-transcriptional modifications of proteins expressed by said cells. In certain embodiments of the present invention, the cellular counterpart of the eukaryotic (e.g., non-primate) host cell is fully functional, i.e., has not been, e.g., inactivated by mutation. Rather the transgenic sequence (e.g., primate) is expressed in addition to its cellular counterpart (e.g., non-primate).

Transfection according to the present invention is the introduction of a nucleic acid into a recipient eukaryotic cell, such as, but not limited to, by electroporation, lipofection, generally via a non- viral vector (vector mediated transfection) or via chemical means including those involving polycationic lipids. Non vector mediated transfection includes, for example, the direct introduction of an isolated TEP siRNAs into a cell. In a transiently transfected cell the, e.g., siRNA only remains transiently. In the context of the present invention there may by a first transfection with at least one nucleic acid molecule with a sequence encoding a transgene expression processing (TEP) protein or TEP functional RNA or, alternatively, directly with a TEP functional RNA (e.g., a siRNA) and a second, subsequent, transfection with a nucleic acid encoding the transgene. Both the first and the second transfection can be repeated. The, e.g., siRNA is introduced during the first transfection, acts, in particular inhibits, a recombination protein (a protein that is involved in the recombination events in the transfected cell). After this the transgene is introduced during the second subsequent transfection.

Transcription means the synthesis of RNA from a DNA template. "Transcriptionally active" refers to, e.g., a transgene that is being transcribed. **Translation** is the process by which RNA makes protein.

An enhancement of secretion is measured relative to a value obtained from a control cell that does not comprise the respective transgenic sequence. Any statistically significant enhancement relative to the value of a control qualifies as a **promotion**.

A **selection marker**, is a nucleic acid that contains a gene whose product confers resistance to an selection agent antibiotic (e.g., chloramphenicol, ampicillin, gentamycin, streptomycin, tetracyclin, kanamycin, neomycin, puromycin) or the ability to grow on selective media (e.g., DHFR (dihydrofolate reductase).

The class of proteins known as **chaperones** has been defined as a protein that binds to and stabilizes an otherwise unstable conformer of another protein and, by controlled binding and release, facilitates its correct fate in vivo, be it folding, oligomeric assembly, transport to a particular subcellular compartment, or disposal by degradation. BiP (also known as GRP78, Ig heavy chain binding protein and Kar2p in yeast) is an abundant about 70 kDa chaperone of the hsp 70 family, resident in the endoplasmic reticulum (ER), which amongst other functions, serves to assist in transport in the secretory system and fold proteins. Protein disulphide isomerase (PDI) is a chaperone protein, resident in the ER that is involved in the catalysis of disulphide bond formation during the post-translational processing of proteins.

CELLULAR METABOLIC ENGINEERING

In cellular metabolic engineering, e.g., the processes inherent in the expressing cell are altered. For example, certain proteins of the secretion pathway are, e.g., overexpressed. Alternatively, recombination events are altered by influencing recombination pathways.

THE PROTEIN SECRETION PATHWAY

The secretion of proteins is a process common to organisms of all three kingdoms. This complex secretion pathway requires most notably the protein translocation from the cytosol across the cytoplasmic membrane of the cell. Multiple steps and a variety of factors are required to for the protein to reach its final destination. In mammalian cells, this secretion pathway involves two major macromolecular assemblies, the signal recognition particle (SRP) and the secretory complex (Sec-complex or translocon). The SRP is composed of six proteins with masses of 9, 14, 19, 54, 68 and 72 kDa and a 7S RNA and the translocon is a donut shaped particle composed of Sec61αβγ, Sec62 and Sec63. Accession numbers (in parenthesis) for the human version of some of these proteins are as follows: hSRP14 (Acc. No. X73459.1); hSRP9 (NM_001130440); hSRP54 (NM_003136) ; hSRPRα (NM_003139); hSRPRβ (NM_021203); hSEC61α1 (NM_013336) ; hSEC61 β (L25085.1) ; hSEC61 γ (AK311845.1).

The first step in protein secretion depends on the signal peptides, which comprises a specific peptide sequence at the amino-terminus of the polypeptide that mediates translocation of nascent protein across the membrane and into the lumen of the endoplasmic reticulum (ER). During this step, the signal peptide that emerges from the leading translating ribosome interacts with the subunit of the SRP particle that recognizes the signal peptide, namely, SRP54. The SRP binding to the signal peptide blocks further elongation of the nascent polypeptide resulting in translation arrest. The SRP9 and -14 proteins are required for the elongation arrest (Walter and Blobel 1981). In a second step, the ribosome-nascent polypeptide-SRP complex is docked to the ER membrane through interaction of SRP54 with the SRP receptor (SR) (Gilmore, Blobel et al. 1982; Pool, Stumm et al. 2002). The SR is a heterodimeric complex containing two proteins, SRα and SRβ that exhibit GTPase activity (Gilmore, Walter et al. 1982). The interaction of SR with SRP54 depends on the binding of GTP (Connolly, Rapiejko et al. 1991). The SR coordinates the release of SRP from the ribosome-nascent polypeptide complex and the association of the exit site of the ribosome with the Sec61 complex (translocon). The growing nascent polypeptide enters the ER through the translocon channel and translation resumes at its normal speed. T*he ribosome* stays bound on the cytoplasmic face of the translocon until translation is completed. In addition to ribosomes, translocons are closely associated with ribophorin on the cytoplasmic face and with chaperones, such as calreticulin and calnexin, and protein disulfide isomerases (PDI) and oligosaccharyl transferase on the luminal face. After extrusion of the growing nascent polypeptide into the lumen of the ER, the signal peptide is cleaved from the pre-protein by an enzyme called a signal peptidase, thereby releasing the mature protein into the ER. Following post-translational modification, correct folding and multimerization, proteins leave the ER and migrate to the Golgi apparatus and then to secretory vesicles. Fusion of the secretory vesicles with the plasma membrane releases the content of the vesicles in the extracellular environment.

Remarkably, secreted proteins have evolved with particular signal sequences that are well suited for their own translocation across the cell membrane. The various sequences found as distinct signal peptides might interact in unique ways with the secretion apparatus. Signal sequences are predominantly hydrophobic in nature, a feature which may be involved in directing the nascent peptide to the secretory proteins. In addition to a hydrophobic stretch of amino acids, a number of common sequence features are shared by the majority of mammalian secretion signals. Different signal peptides vary in the efficiency with which they direct secretion of heterologous proteins, but several secretion signal peptides (i.e. those of interleukin-, immunoglobulin-, histocompatibility receptor-signal sequence, etc) have been identified which

may be used to direct the secretion of heterologous recombinant proteins. Despite similarities, these sequences are not optimal for promoting efficient secretion of some proteins that are difficult to express, because the native signal peptide may not function correctly out of the native context, or because of differences linked to the host cell or to the secretion process. The choice of an appropriate signal sequence for the efficient secretion of a heterologous protein may be further complicated by the interaction of sequences within the cleaved signal peptide with other parts of the mature protein (Johansson, Nilsson et al. 1993).

THE RECOMBINATION PATHWAYS

The recombination pathways, also known as DNA recombination pathways, are cellular pathways that lead to DNA damage repair, such as the joining of DNA molecule extremities after chromosomal double-strand breaks, and to the exchange or fusion of DNA sequences between chromosomal and non-chromosomal DNA molecules, such as e.g. the crossing-over of chromosomes at meiosis or the rearrangement of immunoglobulin genes in lymphocytic cells. The three main recombination pathways are the homologous recombination pathway (HR), the non-homologous end-joining pathway (NHEJ) and the microhomology-mediated end-joining (MMEJ) and alternative end-joining (Alt-EJ) pathway.

THE MECHANISMS OF HOMOLOGOUS RECOMBINATION (HR), NON-HOMOLOGOUS END-JOINING (NHEJ) AND MICROHOMOLOGY MEDIATED END JOINING (MMEJ)

Transgenes use the recombination machineries to integrate at a double strand break into the host genome.

Double-strand breaks (DSBs) are the biologically most deleterious type of genomic damage potentially leading to cell death or a wide variety of genetic rearrangements. Accurate repair is essential for the successful maintenance and propagation of the genetic information.

There are two major DSB repair mechanisms: non-homologous end-joining (NHEJ) and homologous recombination (HR). A third mechanism, called microhomology-mediated end joining (MMEJ) often takes effect when the two major DSB repair mechanisms fail. Homologous recombination is a process for genetic exchange between DNA sequences that share homology and is operative predominantly during the S/G2 phases of the cell cycle, while NHEJ simply pieces together two broken DNA ends, usually with no sequence homology, and it functions in all phases of the cell cycle but is of particular importance during G0-G1 and early S-phase of mitotic cells (Wong and Capecchi, 1985; Delacote and Lopez, 2008). In vertebrates, HR, NHEJ

and MMEJ differentially contribute to DSB repair, depending on the nature of the DSB and the phase of the cell cycle (Takata et al., 1998).

NHEJ: basic mechanisms

Conceptually, the molecular mechanism of the NHEJ process seems to be simple: 1) a set of enzymes capture the broken DNA molecule, 2) a molecular bridge that brings the two DNA ends together is formed and 3) the broken molecules are re-ligated. To perform such reactions, the NHEJ machinery in mammalian cells involves two protein complexes, the heterodimer Ku80/Ku70 associated with DNA-PKcs (catalytic subunit of DNA-dependent protein kinase) and DNA ligase IV with its co-factor XRCC4 (X-ray-complementing Chinese hamster gene 4) and many protein factors, such as Artemis and XLF (XRCC4-like factor; or Cernunnos) (Delacôte et al., 2002). NHEJ is frequently considered as the error-prone DSB repair because it simply pieces together two broken DNA ends, usually with no sequence homology and it generates small insertions and deletions (Moore and Haber, 1996; Wilson et al., 1999). NHEJ provides a mechanism for the repair of DSBs throughout the cell cycle, but is of particular importance during G0-G1 and early S-phase of mitotic cells (Takata et al., 1998; Delacote and Lopez, 2008). The repair of DSBs by NHEJ is observed in organisms ranging from bacteria to mammals, indicating that it has been conserved during evolution.

After DSB formation the key step in NHEJ repair pathway is the physical juxtaposition of the broken DNA ends. NHEJ is initiated by the association of the Ku70/80 heterodimer protein complex to both ends of the broken DNA molecule to capture, tether the ends together and create a scaffold for the assembly of the other NHEJ key factors. The DNA-bound Ku heterodimer complex recruits DNA-PKcs to the DSB, a 460kDa protein belonging to the PIKK (phosphoinositide 3-kinase-like family of protein kinases) (Gottlieb and Jackson, 1993) and activates its serine/threonine kinase function (Yaneva et al., 1997). Two DNA-PKcs molecules interact together across the DSB, thus forming a molecular bridge between both broken DNA ends and inhibit their degradation (DeFazio et al., 2002). Then, DNA ends can be directly ligated, although the majority of termini generated from DSB have to be properly processed prior to ligation (Nikjoo et al., 1998). Depending of the nature of the break, the action of different combinations of processing enzymes may be required to generate compatible overhangs, by filling gaps, removing damaged DNA or secondary structures surrounding the break. This step in the NHEJ process is considered to be responsible for the occasional loss of nucleotides associated with NHEJ repair. One key end-processing enzyme in mammalian NHEJ is Artemis, a member of the metallo-β-lactamase superfamily of enzymes, which was discovered as the

mutated gene in the majority of radiosensitive severe combined immunodeficiency (SCID) patients (Moshous et al., 2001). Artemis has both a 5'→3' exonuclease activity and a DNA-PKcs-dependent endonuclease activity towards DNA-containing ds-ss transitions and DNA hairpins (Ma et al., 2002). Its activity is also regulated by ATM. Thus, Artemis seems likely to be involved in multiple DNA-damage responses. However, only a subset of DNA lesions seem to be repaired by Artemis, as no major defect in DSB repair were observed in Artemis-lacking cells (Wang et al., 2005, Darroudi et al., 2007).

DNA gaps must be filled in to enable the repair. Addition of nucleotides to a DSB is restricted to polymerases μ and λ (Lee et al., 2004; Capp et al., 2007). By interaction with XRCC4, polynucleotide kinase (PNK) is also recruited to DNA ends to permit both DNA polymerization and ligation (Koch et al., 2004). Finally, NHEJ is completed by ligation of the DNA ends, a step carried out by a complex containing XRCC4, DNA ligase IV and XLF (Grawunder et al., 1997). Other ligases can partially substitute DNA ligase IV, because NHEJ can occur in the absence of XRCC4 and Ligase IV (Yan et al., 2008). Furthermore, studies showed that XRCC4 and Ligase IV do not have roles outside of NHEJ, whereas in contrast, KU acts in other processes such as transcription, apoptosis, and responses to microenvironment (Monferran et al., 2004; Müller et al., 2005; Downs and Jackson, 2004).

The NHEJ may be decreased or shut down in different ways, many of which directly affect the above referenced proteins (e.g., the heterodimer Ku80/Ku70, DNA-PKcs, but in particular DNA ligase IV, XRCC4, Artemis and XLF (XRCC4-like factor; or Cernunnos), PIKK (phosphoinositide 3-kinase-like family of protein kinases).

HR: basic mechanisms

Homologous recombination (HR) is a very accurate repair mechanism. A homologous chromatid serves as a template for the repair of the broken strand. HR takes place during the S and G2 phases of the cell cycle, when the sister chromatids are available. Classical HR is mainly characterized by three steps: 1) resection of the 5' of the broken ends, 2) strand invasion and exchange with a homologous DNA duplex, and 3) resolution of recombination intermediates. Different pathways can complete DSB repair, depending on the ability to perform strand invasion, and include the synthesis-dependent strand-annealing (SDSA) pathway, the classical double-strand break repair (DSBR) (Szostak et al, 1983), the break-induced replication (BIR), and, alternatively, the single-strand annealing (SSA) pathway. All HR mechanisms are interconnected and share many enzymatic steps.

The first step of all HR reactions corresponds to the resection of the 5'-ended broken DNA strand by nucleases with the help of the MRN complex (MRE11, RAD50, NBN (previously NBS1, for Nijmegen breakage syndrome 1)) and CtIP (CtBP-interacting protein) (Sun et al., 1991; White and Haber, 1990). The resulting generation of a 3' single-stranded DSB is able to search for a homologous sequence. The invasion of the homologous duplex is performed by a nucleofilament composed of the 3'ss-DNA coated with the RAD51 recombinase protein (Benson et al., 1994). The requirement of the replication protein A (RPA), an heterotrimeric ssDNA-binding protein, involved in DNA metabolic processes linked to ssDNA in eukaryotes (Wold, 1997), is necessary for the assembly of the RAD51-filament (Song and sung, 2000). Then RAD51 interacts with RAD52, which has a ring-like structure (Shen et al., 1996) to displace RPA molecules and facilitate RAD51 loading (Song and sung, 2000). Rad52 is important for recombination processes in yeast (Symington, 2002). However, in vertebrates, BRCA2 (breast cancer type 2 susceptibility protein) rather than RAD52 seems to play an important role in strand invasion and exchange (Davies and Pellegrini, 2007; Esashi et al., 2007). RAD51/RAD52 interaction is stabilized by the binding of RAD54. RAD54 plays also a role in the maturation of recombination intermediates after D-loop formation (Bugreev et al., 2007). In the other hand, BRCA1 (breast cancer 1) interacts with BARD1 (BRCA1 associated RING domain 1) and BACH1 (BTB and CNC homology 1) to perform ligase and helicase DSB repair activity, respectively (Greenberg et al., 2006). BRCA1 also interacts with CtIP in a CDK-dependent manner and undergoes ubiquitination in response to DNA damage (Limbo et al., 2007). As a consequence, BRCA1, CtIP and the MRN complex play a role in the activation of HR-mediated repair of DNA in the S and G2 phases of the cell cycle.

The invasion of the nucleofilament results in the formation of a heteroduplex called displacement-loop (D-loop) and involves the displacement of one strand of the duplex by the invasive strand and the pairing with the other. Then, several HR pathways can complete the repair, using the homologous sequence as template to replace the sequence surrounding the DSB. Depending of the mechanism used, reciprocal exchanges (crossovers) between the homologous template and the broken DNA molecule may be or may not be associated to HR repair. Crossovers may have important genetic consequences, such as genome rearrangements or loss of heterozygosity.

The five Rad51 paralogs are also involved in homologous recombination: Xrcc2, Xrcc3, Rad51B, Rad51C, Rad51D (Suwaki et al., 2011). Rad51 paralogs form two types of complexes: one termed BCDX2 comprises Rad51B, Rad51C, Rad51D and Xrcc2; the other contains Rad51C and Xrcc3 (CX3) (Masson et al., 2001). The first complex has been proposed to

participate in the formation and/or stabilization of the Rad51-DNA complex (Masson et al., 2001). The role of the second complex seems to be branch migration and resolution of the Holliday junction (Liu et al., 2007).

As previously reported, increasing the HR relative to the NHEJ (see US patent pub. 20120231449, which is incorporated herein by reference in its entirety) can be used to enhance and/or facilitate transgene expression.

The present invention focuses on decreasing or shutting down HR. The HR may be decreased or shut down in different ways, many of which directly affect the above referenced proteins (e.g., proteins of the MRN complex (MRE11, RAD50, NBN (previously NBS1, for Nijmegen breakage syndrome 1)) and CtIP (CtBP-interacting protein), RAD51, the replication protein A (RPA), Rad52, BRCA2 (breast cancer type 2 susceptibility protein), RAD54, BRCA1 (breast cancer 1) interacts with BARD1 (BRCA1 associated RING domain 1), BACH1 (BTB and CNC homology 1)). The present invention focuses on the production of RNAs, such as siRNAs to accomplish this goal.

Microhomology-mediated end joining (MMEJ)

When the other recombination pathways fail or are not active, DSBs can be repaired by another, error-prone repair mechanism called microhomology-mediated end joining (MMEJ). This pathway is still needs to be fully characterized and is sometimes also referred to as alternative end-joining (alt-EJ), although it is unclear whether these two processes are based on the same mechanism. The most characteristic feature of this pathway, which distinguishes it from NHEJ, is the use of 5-25 bp microhomologies during the alignment of broken DNA strands (McVey and Lee, 2008).

MMEJ can occur at any time of the cell cycle and is independent of core NHEJ and HR factors, i.e. Ku70, Ligase IV and Rad52 genes (Boboila *et al.*, 2010; Yu and McVey, 2010; Lee and Lee, 2007; Ma *et al.*, 2003). Instead MMEJ initiation relies on its own set of proteins, the most important ones being the components of the MRN complex (MRX in yeast) comprising Mre11, Rad50 and Nbs1 (Xrs2 in yeast), also implicated in the first steps of HR (Ma *et al.*, 2003). Apart from the MRN complex many other factors have been proposed to participate in MMEJ, e.g. CTBP-interacting protein (CtIP; Yun and Hiom, 2009), poly (ADP-ribose) polymerase 1 (PARP1), the ligase III/ Xrcc1 complex, ligase I (Audebert *et al.*, 2004), DNA polymerase θ (Yu and McVey, 2010), and the ERCC1/XPF complex (Ma *et al.*, 2003). However, many more proteins are take part in is process.

It has been suggested that in the absence of other DNA-end binding proteins (like Ku or Rad51) the DSBs are recognized by PARP1 which then initiates their repair through MMEJ (McVey and Lee, 2008). The repair process, similarly to HR, starts with 5' to 3' end resection, which exposes short regions of homology on each side of the break. This processing step is conducted by the MRN complex and regulated by CtIP (Mladenov and Iliakis, 2011). The complementary regions (present in the 3' ssDNA fragments) pair together and the non-complementary segments (flaps) are removed (Yu and McVey, 2010), probably by the ERCC1/XPF complex. Gaps (if any) are then filled in by a polymerase (e.g. DNA polymerase θ or δ (Yu and McVey, 2010; Lee and Lee, 2007)) and breaks joined by the ligase I or ligase III/Xrcc1 complex.

In the absence of immediate microhomology regions at the DNA ends, which is most often the case, a more distant fragment of the repaired molecule can be copied using an accurate DNA polymerase (e.g. polymerase θ). This duplicated region then participates in the alignment of DNA ends, which results in an insertion in the created junction. This more complex variant of microhomology-mediated repair has been termed synthesis-dependent MMEJ (SD-MMEJ) (Yu and McVey, 2010).

Although MMEJ was thought to act as an alternative recombination repair pathway, it has been shown to be very efficient in the process of IgH class switch recombination in B lymphocytes (Boboila *et al.*, 2010), suggesting that it might be more than a backup mechanism. It is also possible that some DSBs, e.g. incompatible overhangs or blunt ends (which are poor NHEJ and/or HR targets) might be more efficiently repaired by MMEJ (Zhang and Paull, 2005).

TABLE D lists some of the key genes in each of the three pathways, which are therefore also key targets for influencing each of the three pathways (see also US Patent Publication 20120231449, which is incorporated herein by reference in its entirety). Also included in the table are DNA repair proteins such as MDC1 and MHS2. MDC1 is required to activate the intra-S phase and G2/M phase cell cycle checkpoints in response to DNA damage. However, MDC1 also functions in Rad51-mediated homologous recombination by retaining Rad51 in chromatin.

"Knock-down" in the context of the present invention conveys that expression of the target gene is reduced, such as by 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more. Complete knock-down means that there is no detectable expression of the target gene anymore. TABLE D shows also the results obtained with certain knock-down targets. As the person skilled in the art will appreciate there are variations in the nucleic acid sequences of the

targets so that variants of the genes, in particuar variants that display a sequence identity of 80%, 90% or 95% are part of the present invention.

PROTEIN PROCESSING AND METABOLIC PROTEINS

This category of proteins that can be used for cellular metabolic engineering neither belongs to the protein secretion pathway nor the recombination pathway but otherwise influence processes inherent in the expressing cell.

The protein processing or metabolic proteins are often enzymes such as chaperones (see defintions of chaperons above), proteins isomerases, sugar adding enzymes (e.g. sialyl or glycosyl transferases) or phosphatases, or control the cell energy level or mitochondrial function.

TABLE A sets forth a list of proteins that have been expressed (exp) and/or whose expression has been "knocked-down" (KD) under the subheading "protein processing and metabolic proteins".

VECTOR DESIGN

Among non-viral vectors, transposons are particularly attractive because of their ability to integrate single copies of DNA sequences with high frequency at multiple loci within the host genome. Unlike viral vectors, some transposons were reported not to integrate preferentially close to cellular genes, and they are thus less likely to introduce deleterious mutations. Moreover, transposons are readily produced and handled, comprising generally of a transposon donor plasmid containing the cargo DNA flanked by inverted repeat sequences and of a transposase-expressing helper plasmid or mRNA. Several transposon systems were developed to mobilize DNA in a variety of cell lines without interfering with endogenous transposon copies. For instance, the PiggyBac (PB) transposon originally isolated from the cabbage looper moth efficiently transposes cargo DNA into a variety of mammalian cells.

Epigenetic regulatory elements can be used to protect the cargo DNA from unwanted epigenetic effects when placed near the transgene on plasmid vectors. For example, elements called matrix attachment region (MARs) were proposed to increase cargo DNA genomic integration and transcription while preventing heterochromatin silencing, as exemplified by the potent human MAR 1-68. They can also act as insulators and thereby prevent the activation of

neighboring cellular genes. MAR elements have thus been used to mediate high and sustained expression in the context of plasmid or viral vectors.

As shown herein, with the proper vector design, the favorable properties of epigenetic regulators, in particular MAR elements, may be combined with those of transposable vectors.

Transposons and transposon based vectors of the present invention can be used in cellular metabolic engineering, for instance to express secretion proteins of different secretion pathways described herein. They are also particularly useful when multiple rounds of cargo DNA introduction are required. This was confirmed when testing multiple proteins of the cell's secretory pathway, where the transfection of multiple vectors and/or multiple successive transfection cycles may exhaust available antibiotic or other selection methods. The ability to quickly express therapeutic proteins without a need for antibiotic selection is also of particular interest, for instance when multiple therapeutic protein candidates must be expressed for screening purposes, since significant amounts of proteins can be obtained from unselected cell populations 2-3 weeks after transfection. In particular, MAR-containing transposon vectors are thus a promising addition to the currently available arsenal of expression vectors.

The experimental approaches chosen, allowed (as opposed to approaches that rely on antibiotic based assays), a distinction between effects based on (1) cargo DNA copy number and effects based on (2) cargo DNA expression levels.

MAR 1-68 was particularly efficient, when located centrally between the ITRs of the of a PiggyBac transposon as it did not decrease transposition efficiency. MAR X-29 also worked well at the edges of the transposon without decreasing transposition efficiency or expression.

Interestingly, the extent of the MAR-mediated activation of transposed genes was reduced when compared to that of spontaneous plasmid integration. Furthermore, the level of expression, when normalized to, e.g., transgene copies, was higher from the transposons than those obtained from the spontaneous integration of the plasmids in the absence of the transposase. This effect was observed irrespective of the size of the constructs, of the presence of the MAR or of promoter strength. This would be expected if transposition might often occur at genomic loci that are relatively permissive for expression, for instance because open chromatin structures may be more accessible to both the transposase and transcription factors. In this

respect, previous studies have suggested that transposons may preferentially integrate within gene introns, at promoters, or at genomic loci with lower propensity for silencing, although this has remained a matter of debate. Alternatively, the co-integration of many plasmid copies at the same genomic locus, as elicited by spontaneous integration events, may lead to the formation of heterochromatin and to the silencing of repetitive sequences, which the MAR would oppose, whereas single-copy transposon integration may be less prone to such chromatin-mediated silencing. In addition, the integration of transposons at multiple independent genomic loci makes it likely that at least one copy landed in a favorable genomic environment and is expressed, whereas plasmid integration was found to occur predominantly at just one genomic locus.

The highest expression levels per cargo DNA, e.g., transgene/TER were obtained from a MAR-containing transposon when coupled to a strong promoter. It was surprising to find high expression levels could be obtained from a few transposed cargo DNA copies, e.g., not more than 20, 15, 10, or 5. If high productivities can nevertheless be obtained, fewer integrated, e.g., cargo DNA copies are advantageous, as it decreases the probability of point mutation occurrence in one or in a subset of the transgenes, as elicited from spontaneous mutagenic events. In addition, transposase-mediated integration events are less mutagenic than the DNA repair and recombination mechanisms involved in spontaneous plasmid integration, which can lead to incomplete or rearranged transgene copies.

The high efficiency of genomic integration by the *piggyBac* (PB) transposon is also be favorable when the amount of target cells is limiting, for instance for the non-viral transfer of therapeutic genes into primary stem cells to generate clonal populations for, e.g., cell-based therapies or regenerative medicine. In this context, physiological expression levels from a few transposed cargo DNA copies and the frequent occurrence of transposition events, thus obviating the need for antibiotic selection, is advantageous, since the use of antibiotic resistance genes and/or unreliable, e.g., transgene expression may raise safety concerns.

Effect of MAR inclusion on transposition efficiency

As antibiotic resistance does not necessarily reflect efficient transgene expression, the green fluorescent protein (GFP) expressed from a strong GAPDH cellular promoter derivative was used as an indicator. To test whether adding a MAR element to the PB transposon may affect transposition efficiency and transgene expression, and to assess whether the location of the MAR in the construct had any influence on these effects, a series of transposon donor

constructs were designed containing the GFP and puromycin resistance (Puro) gene, in which the MAR 1-68 or a control neutral spacer DNA sequence were inserted at different positions in the plasmid **(Figure 1)**. The parental Puro-GFP transposon plasmid without an insert was used as a control of transposition, to distinguish the impact of increased transposon size relative to effect of the MAR or spacer sequence addition.

In the presence of the transposase, the highest level of GFP expression from unselected cells was observed when the MAR was centrally located, but not when the MAR was placed downstream of the GFP coding sequence, nor when inserted outside of the transposed sequence as expected **(Figure 3A)**. In the presence of puromycin selection, the MAR-mediated activation was reduced, either with or without the transposase, while the GFP expression averages were increased by one order of magnitude **(Figure 3B)**. This confirmed that puromycin selection yielded only the minority of the cells that display the highest expression levels, as proposed above from the quantitation of transposition events. It further indicated that the transposable vectors containing a centrally located MAR yielded similar expression levels when compared to their plasmid counterpart transfected without the transposase.

Effect of MAR inclusion on the copy number of integrated transposon

Higher GFP fluorescence levels may result from an increased transcription of the transgenes and/or by the integration of more transgene copies. This was assessed by quantifying the number of genome-integrated transgene copies resulting from the various types of vectors. Total genomic DNA was isolated from pooled populations of cells, either after cytofluorometric sorting of fluorescent cells from unselected populations or after selection for *puromycin* resistance. The transgene copy number was determined by quantitative polymerase chain reaction (qPCR) analysis of the GFP coding sequence relative to the cellular β2-microglobulin (B2M) gene. In the absence of antibiotic selection, the average number of transgenes integrated by either the transposase or by cellular recombination enzymes were similar, around 1-6 copies per genome, and they were not significantly affected by the MAR or control sequence **(Fig. 4A)**. However, the lowest copy number was obtained when the MAR was included at the transposon edge, supporting our earlier conclusion that it decreases transposition at this location. After selection for highly expressing cells with puromycin, the number of transposed transgenes was in a similar 2-7 copy range **(Fig. 4B)**. However, the number of transgenes copies integrated in the absence of the transposase was generally significantly higher, ranging from 6 to 14 copies. This can be readily explained by the fact that spontaneous integration usually results in the integration of concatemers of multiple plasmid copies at a single genomic locus (results not

shown), and that higher transgene copy numbers should lead to higher expression levels when cells subjected to silencing effects have been removed by antibiotic selection. Taken together with the prior conclusion that antibiotic selection preferentially yields highly expressing cells, this also indicated that spontaneous plasmid integration results in a more variable number of transgene copies than transposable vectors.

GFP expression was then normalized to the gene copy number to assess the intrinsic expression potential of the vectors, independently from their propensity to integrate in the genome. Overall, lower expression per transgene copy was obtained from unselected cells, or from antibiotic-selected cells transfected without transposase or centrally-located MAR, indicating that transgene expression is influenced both by the inclusion of the epigenetic regulatory element and by the mode of transgene integration **(Fig. 5)**. Expression per gene copy was generally increased by the transposase, when assessed from various vectors and combination of elements, and this was observed with or without antibiotic selection. The highest levels of expression per transgene copy were obtained after antibiotic selection from the cells generated with the transposon vector containing the MAR element centrally located and in presence of the transposase. Inclusion of the MAR immediately downstream of the GFP coding sequence did not increase transgene expression significantly, as noted earlier for the absolute levels of expression.

Finally, it was assessed whether the favorable effect of MAR 1-68 on expression may be specific to the strong human GAPDH promoter used here, or whether it would also occur with other promoters. Thus we replaced the human GAPDH promoter driving GFP expression by the weaker simian virus 40 (SV40) early promoter. Use of the weaker promoter yielded comparable numbers of GFP-positive cells and of integrated transgenes, indicating that the transposition efficiency is not altered by transgene expression (results not shown and **Fig. 2A and Fig. 4B**). However, the absolute levels of expression were lower with the SV40 promoter (not shown vs. **Fig. 3B**). In addition, expression normalized to the transposon copy number was decreased by 4.6-fold by the use of the SV40 promoter in the absence of the MAR, and by 3.1-fold with MAR 1-68 (results not shown). This indicated that the MAR could partially, but not fully prevent the decrease of expression resulting from the use of a weaker promoter, even in presence of the transposase. Overall, it could be shown that a few integrated copies are sufficient to obtain high transgene expression from transposons, and that the highest expression per transgene is obtained when, in this context, MAR-68, is placed upstream of the strong promoter.

CHO-M cells were electroporated once or twice with a single transgene MAR X_29-containing transposable vector. Transposition efficiency was highest after electroporation (30%-45% of

the cells showed stable expression). However, transgene expression levels were similar to chemical transfection, which showed lower positve cells, ergo lower transposition efficiency. **Fig. 7** shows the results with light and heavy chains of therapeutic Immunoglobulins inserted upstream of the MAR X_29 and titles ranging from 1 to 8 µg/ml were obtained. The levels were further increased to 23-55 µg/ml by sorting the expressing cells (**Fig. 7C**).

Expression of transgenes can also be substantially increased, often independent of the use of transposons by specific vector designs, in particular by the use of specific MAR element(s) at specific locations relative to the transgene and, preferably a combination of those MAR element(s) with promoters, enhances or fusions thereof.

A respective vector may contain MARs that flank the transgene expression cassette. For example, the vector may contain, e.g., upstream MARs (one or more) and downstream MARs (one or more), e.g., one MAR positioned upstream and one MAR positioned downstream of a transgene expression cassette **(Fig. 18A, Fig. 18B, Fig. 19**). The vector may contain an integrated puromycin resistance gene under the control of the SV40 promoter. The transgene may be under the control of the human GAPDH promoter fused to the human cytomegalovirus (CMV) immediate-early genes enhancer (in particular the CGAPD fusion promoter as discussed above).

The highest percentile of high and very high producer cells (% M3/M2), as assessed for GFP fluorescence by FACS analysis and the least variability, could be obtained using are 1_6R2, 1_68R2 and X_29R3 as the upstream MAR (over 80%, 80% and over 80%). Thus, a percentile of high and very high producer cells (% M3/M2), of more than 70%, more than 75%, or more than 80%, are well within the scope of the present invention. As the person skilled in the art will understand, certain deviation from the specific sequence of theses MARs are permissible. Accordingly, vectors containing nucleic acid sequences having more than 80%, 85%, 90%, 95% sequence identities with SEQ ID Nos: 6, 7, 8, 9 and 10 are within the scope of the present invention (**Fig. 18A, Fig. 18B**).

Loss of expression in the bioreactor and/or in the absence of selection pressure often limits recovery of the protein of interest. Vectors containing the 1_68R2, 1_6R2 and X_29R3 MAR derivatives as the upstream MAR were tested over a period of 5 weeks of culture without selection, and GFP fluorescence was assessed weekly over this period. When considering the

percentile of the M3 subpopulation, it was found that the 1_6R2 element as an upstream MAR and the unrearranged MAR 1-68 as a downstream MAR were the best tested combination in vectors with at least one upstream and one downstream two MARs (well above 80% after more than 2, 3, 4 weeks) (see, **Fig. 19**).

A similarly designed vector may also contain, e.g., just downstream MARs (one or more), e.g., one MAR positioned downstream of a transgenes expression cassette (**Figs. 20A, 20B, 21 and 22**) and no upstream MAR. The vector may also in this case contain an integrated puromycin resistance gene under the control of the SV40 promoter. The transgene may be under the control of the human GAPDH promoter fused to the human cytomegalovirus (CMV) immediate-early genes enhancer. See, e.g., SEQ ID NO: 11 are others having sequence identies of more than 80%, 85%, 90% or 95%. Excellent results were achieved in such a single MAR constellation with X_29 as a MAR.The percentile of high GFP expressing cells (determined as above) and also the stability of expression over time (determined as above) is better then, e.g., that of high performing vectors in which MARs flanked the transgene expression cassette, namely a vector comprising a MAR 1_6R2 upstream and an unrearranged MAR 1-68 downstream (See **Fig. 22**). This finding contrast the well established assumptions that MARs are most effective when they flank the transgene (see US patent 5,731,178). Stability of expression means that a DNA of interest, e.g., a transgene, is expressed by a cell population even after a certain period of time, e.g., after more than 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 weeks at rate comparable (not more than 20%, 10 or 5% less) and particularly even higher as up to two weeks of the commencement of expression. Often stable expression is associated with a high percentile (e.g. more than 80%) or highly expressing subpopulation of cells.

CELLULAR METABOLIC ENGINEERING: **SECRETION PROTEINS**

The secretion of heterologous proteins such as IgGs id dwarfted by improper polypeptide processing and low IgG production in cultured cells such as CHO cells.

It was observed that the expression of stress-induced chaperones like BiP is induced and that the chaperone is correctly localized in the ER and capable of interacting with the IgG precursor chains. However, IgGs containing particular variable sequences such as those found in Infliximab are nevertheless incorrectly processed and assembled, which leads to poor secretion. Therefore, the activation of the UPR response in these cells remains ineffective in rescuing significant level of immunoglobulin.

SRP14 was shown to be implicated in a molecular step of the secretion pathway that is limiting in CHO cells over-expressing an exogenous protein. Interestingly, this limiting step also occurs for the easy-to-express Trastuzumab that readily leads to high expressor clones. This conclusion followed the finding that expression of the human SRP14 readily restored expression of the LP clones, but that it also increased the secretion of the easily expressed IgG. SRP14 expression was found to increase the processing and availability of LC and HC precursors and to yield comparable levels of secretion for both types of IgGs. Overall, it demonstrated that SRP14 may be generally limiting when secreted proteins such as IgGs are over-expressed in CHO cells.

The strong effect obtained from the expression of SRP14 in CHO cells as observed in this study was unexpected and suggested that SRP14 causes an extended delay of the LC elongation in the difficult-to-produce IgG producer clones (**Fig. 12B, point 1**).

Prior work indicated that the signal peptide that emerges from a translating ribosome first interacts with the SRP54 subunit of the SRP particle, while association with SRP9 and SRP14 may block further elongation of the nascent polypeptide, resulting in translation arrest (Walter and Blobel 1981). In a second step, the ribosome-nascent polypeptide-SRP complex docks to the ER membrane through its interaction with the SR receptor (Gilmore et al., 1982; Walter et al., 1982). The SR may then coordinate the release of the SRP from the ribosome-nascent polypeptide complex and the association of the exit site of the ribosome with the translocon channel, through which the growing nascent polypeptide enters the ER (Lakkaraju et al., 2008). Then, the translation-coupled translocation may resume, leading to the removal of the signal peptide and to the synthesis of properly processed and secreted polypeptides.

Proper processing of the difficult-to-express IgGs might require an unusually long translational pausing, if the kinetics of docking onto the ER may be slower for particular combinations of IgG variable domain and signal peptide sequences, because of unfavorable structures of the nascent peptide. Thus, modulation of the translation arrest kinetic by expression of the exogenous human SRP14 component was considered in turn improve proper ER docking and the translocation of the pre-LC, and thus restore an efficient processing of the signal peptide (**Fig. 12B, point 2**). Consistently, the lowering of SRP14 levels in human cells lead to a lack of translation elongation delay in polysomes, which may result in the overextension of the nascent polypeptides beyond a critical length, after which the SRP may no longer properly target the

The analysis of secretion intermediates and of possible cellular stress responses, followed by the systematic search of the upstream limiting activities that cause such stress response, and then finalized by the engineering of the CHO cell secretion metabolism has lead to a better understanding of the metabolic limitations of these cells and how to address them.

Heterologous expression of SRP14 restores secretion and LC processing

HP (high producer) and LP (Low producer) clones of IgG were co-transfected with a vector encoding the SRP14 component of SRP and with a neomycin resistance plasmid. Individual cells in the neomycin-resistant pool were separated by limiting dilution and subsequently tested for growth and immunoglobulin secretion in shaken culture dish batches. SRP14-expressing LP-derived subclones secreted significantly higher antibody amounts than their parental counterparts throughout the culture, and they yielded similar immunoglobulin titers as the HP SRP14-expressing subclones (**Fig. 8A and 8B, left panels**).

Expression of SRP14 did not affect cell viability, but it appeared to slow down and prolong the growth of HP cell cultures up to similar cell densities (**Fig. 8A and 8B, right panels**). Culture supernatants of the various subclones were collected and analyzed for antibody concentration. As shown in**Fig. 8C**, SRP14 expression enhanced the secretion from LP cells, leading to a 7-fold increase of the IgG specific productivity. Moreover, exogenous expression of SRP14 also improved IgG secretion from the HP subclones, leading to a 30% increase of the specific productivity. Interestingly, individual subclones expressing SRP14 secreted the difficult- and easy-to-express IgGs at essentially identical average rates, with median specific productivities exceeding 30 picogram per cell and per day (pcd). These very high IgG secretion levels were maintained for more than 6 months of culture, indicating that it is a stable property of SRP14-expressing cells.

To further investigate the relationship between SRP14 expression and IgG productivity, the SRP14 mRNA levels of the 5 individual SRP14-expressing LP subclones were analyzed by relative quantitative PCR. As shown in **Fig. 8D**, subclones overexpressed SRP14 at levels that ranged from 50 to nearly 200-fold over that of the endogenous CHO cell SRP14 mRNA. This was accompanied by an IgG secretion enhancement of 4 to 6-fold as compared to the LP control cell clone. Interestingly, the highest specific productivity was obtained from a subclone overexpressing SRP14 at an intermediate level, approximately 100-fold over the CHO cell

secreted protein to the ER. Thus, CHO cells SRP14, and possibly also SRP54, would have a reduced affinity for signal sequences of the heterologous human Infliximab protein, leading to incorrect ER docking and/or to the elongation of the nascent peptide before proper docking has occurred. This would be corrected by the over-expression of human SRP14, lengthening of the time period during which the arrested ribosome-SRP complex may search for a properly organized docking site on the ER, despite the 'molecular jam' of over-expressed IgG proteins occurring at ER gates.

Consistently, overexpression of the SR and translocon, which may increase the ER capacity in terms of translocation, also resulted in an improvement of secretion, even in the absence of human SRP14 overexpression. Finally, it was demonstrated that the metabolic engineering of the secretory pathway, by the co-expression of combinations of human SRP, translocon and SR subunits, leads to further improvement of the protein secretion cellular capacity, yielding even higher secretion levels. Overall, it was concluded that SRP proteins, its receptor and the translocon may be generally limiting when secreted proteins such as human immunoglobulins are over-expressed by CHO cells.

Little has been known about the abundance of SRP and ER membrane components relative to secreted proteins and to ribosomes in different cell types, but translocation defects may conceivably arise in cells expressing high amounts of a recombinant protein. For instance, the SR and/or the translocon may become limiting when secreted proteins are expressed at abnormally high levels, or SRP14 may occur at sub-stoichiometric levels in CHO cells relative to other SRP subunits. Consistently, SRP9 and SRP14 are present in a 20-fold excess over other SRP proteins in primate cells but not in mouse cells, and over-expression of human SRP14 in normal human cells did not increase the efficiency of the secretion of the alkaline phosphatase. Furthermore, the human SRP14 is larger than its rodent counterpart, as it contains an alanine-rich tail at its C-terminus that is not found in the rodent SRP14. Thus, incorporation of the larger human SRP14 in the CHO SRP might lead to the formation of a functional SRP chimera of higher activity, in a dominant-positive effect.
The finding that the expression of cytosolic SRP components such as SRP14 leads to efficient processing and secretion of over-expressed proteins in CHO cells points to a bottleneck that can be used to improve recombinant protein yields. This bottleneck limits the expression of distinct and unrelated IgGs, and possibly also of the numerous other monoclonal antibodies and derivatives that constitute by far the most abundant class of recombinant therapeutic proteins.

endogenous SRP14 mRNA (SRP14-LP subclone E, not shown). This implied an interdependence of the level of SRP14 overexpression and IgG specific productivity up to a threshold level of SRP14 corresponding to a 100-fold increase over that of the endogenous expression level. This suggested that other components of the secretory pathway may in turn become limiting at very high levels of SRP14, and that balanced expression of the pathway components may be required for optimal IgG expression.

To test whether the increased specific productivity obtained during clonal cell line evaluation could be applied to a production process, the best HP and LP SRP14-expressing subclones were tested in shaken cultures dishes in fed-batch conditions (i.e. LP subclone E and HP subclone B of **Fig. 8**). The SRP14-expressing LP subclone yielded similarly high numbers of viable cells and immunoglobulin titers than the SRP14-expressing HP subclone, with a maximum of $8x10^6$ cells/ml and above 2g per liter at the end of the production run **(Fig. 9A and B)**.

The impact of SRP14 overexpression on immunoglobulin synthesis for these two subclones was next tested. This revealed that expression of human SRP14 in the LP-derived subclone led to normally processed and mature LC competent for folding and IgG assembly (**Fig. 10A,** LP lane S vs. lane -). Migration of the free HC was not affected, indicating that SRP14 expression acted specifically on the misprocessed LC of the difficult-to-express protein. Strikingly, SRP14 expression fully abolished the accumulation of aggregated LC in the Triton X-insoluble fraction (Fig. 9A, bottom panel). Expression of the control GFP protein did not improve protein solubility, nor did it restore proper processing of the LC (**Fig. 10A,** lane G of LP cells). Expression of SRP14 had no effect on the HC and LC migration pattern obtained from the HP subclone, and little effect was observed on the amount of the free chains and fully assembled IgG when compared to controls (**Fig. 10A,** lane G of HP cells).

Cycloheximide-based chase assays were performed to investigate the IgG folding and assembly kinetic as well as the fate of the IgG aggregates in the SRP14-expressing Infliximab producer subclone. In contrast to the parental LP cells exhibiting aggregated LC incompetent for IgG assembly, the SRP14-expressing LP subclone no longer accumulated Tx-100 insoluble LC (Fig. 9B, bottom panel). However, the free LC remained in small amounts relative to the free HC, as also noted for the HP cells, indicating that it was quickly incorporated into HC-LC dimers and the mature IgG and that it may be limiting IgG assembly (**Fig. 9A and 9B**). Collectively,

these results implied that SRP14 may play an essential role in LC processing by LP cells, and that additional SRP protein expression could improve the secretion of the difficult-to-express and easy-to-express IgGs up to similar levels.

Engineering of ER translocation improves recombinant IgG secretion

Given that overexpression of heterologous SRP14 increased IgG secretion up to a given threshold, it was reasoned that other components of the secretion pathway that interact directly or indirectly with SRP14 may become limiting in the SRP14-LP subclones. We therefore explored whether the overexpression of other components of the secretion pathway may also improve IgG expression, either alone or in combination with SRP14. These included the human SRP9 and SRP54 proteins that constitute the SRP complex together with SRP14, and subunits of the SRP receptor (SR) and the Sec61α, β and γ subunits of the translocon.

In a first set of experiments, the best performing LP clone, namely LP clone E of Fig1., was transfected with expression vectors encoding SRP proteins or translocon proteins alone or in combinations. The resulting LP polyclonal cell pools were then evaluated for IgG production in batch cultivation. Expression of SRP components or of translocon proteins increased immunoglobulin secretion from these re-transfected LP polyclonal cell pools (**Fig. 11A** and data not shown). Compared to SRP14 expression alone, the overexpression of SRP protein combinations or of the translocon improved the specific productivity of transfected-LP polyclonal cell pools by an additional 20% to 40% (**Fig. 11A**, comparison of median values). These results clearly indicated that particular combinations were more potent to restore Infliximab secretion than the SRP14 expression alone, such as those consisting of the expression of the three SRP polypeptides and its receptor (SR), or the co-expression of the SR and of the translocon (**Fig. 11A**).

Whether the SRP14-expressing LP subclone E could be optimized further by the expression of SR and/or translocon combinations was also assessed. Compared to the 30 pcd of the SRP14-LP subclone E, polyclonal cell pools selected after transfection with the SR proteins and the translocon yielded specific productivities above 60 pcd for the difficult-to-express immunoglobulin (**Fig. 10B**). It was concluded that SR proteins and the translocon expressing vectors can also be used to generate clones with increased specific productivities as compared to SRP14-LP clone E, and that an approach based on series of consecutive transfection and selection cycles may be successfully applied.

CELLULAR METABOLIC ENGINEERING: Knock-down of recombination pathways and expression of TEPs and TEP functional RNA

To influence recombination, crucial DNA recombination genes can be silenced. Targets for knock-down are genes known from literature as crucial for particular recombination repair pathways in mammals (**Table D**). Since most of these genes are necessary for cell survival and development, their permanent silencing could result in reduced cell viability. Therefore, silencing these target genes transiently using RNA interference (RNAi) is preferred.

In cells treated with a mix of siRNAs against factors involved in the first steps of NHEJ, i.e. Ku 70, Ku80 and DNA-PKcs (one siRNA duplex per protein), the frequency of NHEJ events was significantly lower than in the untreated cells. Instead HR was more efficient in these cells, resulting in a reduced NHEJ to HR ratio (from 2.7:1 to 1.3:1). Conversely, in cells treated with siRNA targeted against an essential HR factor - Rad51, the HR-dependent GFP reconstitution was almost completely abolished, which increased the NHEJ:HR ratio to 5:1. These results seem to indicate that the HR and NHEJ reporter assay is sensitive enough to be used in the extrachromosomal form. They also further challenge the reliability of the previously used CHO mutant cell lines (notably the 51D1 cells, originally published as HR-negative cells, but seemingly capable of performing HR according to this assay –data not shown).

Increase of GFP expression and integration in the presence of MAR. In parallel experiments, CHO cells treated with different anti-HR or anti-NHEJ siRNAs were transfected with GFP or MAR-GFP containing plasmids. After two weeks of antibiotic selection the cells were assayed for GFP expression and integration by FACS and qPCR respectively. In all conditions tested the addition of the MAR resulted in a 4-5 fold increase of GFP expression compared to cells transfected with a plasmid without the MAR (**Fig. 13, Fig. 14A**), which is in line with previous reports (Grandjean *et al.*, 2011). This increase was accompanied by an approx. 3-fold increase of integrated GFP copy number in cells transfected with the MAR-GFP plasmid as compared to cells that received the no-MAR vector (**Fig. 13, Fig. 14B**). There was also a 2-fold increase in the average GFP fluorescence per gene copy **(Fig. 13, Fig. 14C**), possibly due to a more favorable localization of the integration site (e.g. in a region rich in euchromatin). Therefore it could be hypothesized that MAR elements possess the ability to direct genes to genomic loci permissive for gene expression. A combination of antiHR siRNAs with MAR (siRNA: RAD51, Rad51C and Brca1) resulted in a fold change in mean GFP

expression of above 11, while a singular antiHR-siRNA (si:RNA: Rad51 lead to a fold change of under 9. Other combinations such as a combiantion of siMMEJ siRNAs with MAR also led to improved expression relative to the singluar siRNA (reualts not shown). Accordingly, combination of different siRNAs (2, 3, 4, 5, 6, 8, 9, 10) from targets in the same or different metabolic pathways, in particular the recombination pathway, are within the scope of the present invention.

No effect of NHEJ gene knock-down on transgene expression and integration. Treatment with siRNAs against the NHEJ proteins did not seem to significantly influence stable transgene expression. There was also no significant change in GFP copy number in the genome compared with the untreated cells (except in cells treated with the anti-53BP1 siRNAs, but only in the absence of the MAR, possibly pointing to an effect unlike recombination by the NHEJ pathway).

Increase of transgene expression and integration in the absence of HR factors. In contrast to the knock-down of NHEJ factors, the presence of siRNAs against HR proteins often resulted in a significant increase of stable GFP expression as compared to the untreated cells (except for Brca2, for which there was a significant decrease, but again only in the absence of the MAR) (**Fig. 15A**). As was the case with the silencing of NHEJ genes, the presence of the MAR resulted in an increased stable GFP expression and integration, as well as the expression per gene copy. This time however the silencing of HR factors enhanced this effect by 5 to 7-fold (the most striking being the knock-down of Rad51 reaching 7.4-fold higher GFP expression levels) (**Fig. 15A**), which could indicate that HR proteins counteract the positive effect of the MAR element on transgene expression. In the absence of the MAR, the increase in GFP expression was correlated with an elevated GFP copy number in the genome (*results not shown*). Surprisingly, this was not the case in the presence of MAR, indicating that the MAR-mediated increase in copy number was not affected by the HR protein knock-down (except for Rad51D siRNA). This seems to suggest that the absence of a functional HR repair pathway does not enhance the number of recombination events promoted by the MAR. Instead, it might stimulate the integration of the MAR and transgene in a more favorable locus allowing for its more efficient expression. This view is also supported by the elevated expression of individual GFP copies in the presence of the MAR in cells treated with the anti-HR siRNAs (Fig. 14C). Another possibility is that the number of plasmid copies integrated into the genome is already at its

maximum in the control cells (with the amount of plasmid DNA used here) and cannot be further increased even in conditions more beneficial for the MAR.

Taken together these results suggest that the process of MAR-mediated transgene integration is preferentially mediated by a pathway opposed to homologous recombination, although likely not NHEJ since knock-down of its components had no effect on integration or expression. It could be hypothesized that this alternative pathway is less active in the presence of a functional HR pathway, but becomes more important if HR disabled.

EXPRESSION OF ANTI-HR shRNAS TO INCREASE THE EXPRESSION OF THERAPEUTIC PROTEINS

Three siRNA targeting Rad51 were converted into shRNA sequences that can form hairpin structures, and the shRNA coding sequence was inserted into a *piggyBac* transposon vector under the control of a GAPDH enhancer and CMV promoter fusion and followed by the MAR X-29, but devoid of an antibiotic selection gene. Suspension-adapted CHO-M cells were transfected three times with the transposon donor plasmid and the transposase expression vector, after which 30 individual cell clones were randomly picked using a ClonePix device. Parental cells as well as the pool of shRNA-expressing cell pool and clones were re-transfected with a GFP expression plasmid (namely the Puro-GFP-MAR X_29 construct), which was followed by puromycin selection of polyclonal pools of GFP-expressing cells. Comparison of the GFP fluorescence profile indicated that a higher proportion of medium to highly fluorescent cells (M2 cell population) or very highly fluorescent cells (M3 cell population) were obtained from the cell pool transfected with the shRNA vector as compared to the parental CHO-M cells **(Fig. 17A)** ~~(Fig. 15A)~~.

Several shRNA-expressing clones mediated very high GFP levels, with over 80% of the antibiotic resistant cells being in the highly fluorescent M3 subpopulation 10 days after transfection, as exemplified by clone 16 and clone 26 (**Fig. 17A and B**). High levels of GFP fluorescence were maintained in these two clones after 35 days of further culture without selection (**Fig. 17C**). In contrast, clone 17 did not express very high levels of GFP at day 10, and GFP expression appeared to be unstable (**Fig. 17B and C**). Intermediate expression levels and stability were obtained from clone 8 and clone 22. These clones were also transfected with expression plasmids encoding the light and the heavy chain of the difficult-to-express Infliximab therapeutic antibody. As before, the clones 16 and 26 produced the highest levels of the

antibody, followed by clones 8 and 22, whereas clone 17 expressed amounts of the immunoglobulin that were similar to the 0.5 to 1 pcd obtained from the parental CHO-M cells (**Fig. 17D** and data not shown). Cell clones displaying Rad51 mRNA levels that were most significantly reduced yielded high expression of both GFP and of the Infliximab immunoglobulin, indicating that increased transgene expression resulted from the decreased expression of the recombination protein (*results not shown*).

Thus, higher and more stable production levels of secreted therapeutic proteins such as Infliximab can be achieved from cells expressing a Rad51-targeting shRNA or from cell transiently transfected by siRNAs such as the Rad51-targeting siRNA.

MATERIALS AND METHODS

Plasmids and DNA vectors

The PB transposase expression vector pCS2+U5V5PBU3 contains the *PB* transposase coding sequence surrounded by the 5' and 3' untranslated terminal regions (UTR) of the *Xenopus laevis* β-globin gene. This plasmid was constructed as follows: the 3' UTR 317 bp fragment from pBSSK/SB10 (kindly provided by Dr S. Ivics) was inserted into pCS2+U5 (Invitrogen/Life Technologies, Paisley, UK) to yield pCS2+U5U3. The PB transposase coding sequence (2067 bp, GenBank accession number: EF587698) was synthesized by ATG:biosynthetic (Merzhausen, Germany) and cloned in the pCS2+U5U3 backbone between the two UTRs. The PB control vector corresponds to the unmodified pCS2+U5 plasmid (Figure 1, left panel).

The different transposons vectors used in this study were generated by introducing the *PB* 235 bp 3' and 310 bp 5' inverted terminal repeats (ITRs), synthesized by ATG:biosynthetic (Merzhausen, Germany), into the pBluescript SK- plasmid (pBSK ITR3'-ITR5', Figure 1, right panel). The puromycin resistance gene ($Puro^R$), under the control of the SV40 promoter from pRc/RSV plasmid (Invitrogen/Life Technologies), was then inserted between the two ITRs. The MAR 1-68 and MAR X-29 elements, the puromycin resistance and GFP genes used in this study were as previously described. The immunoglobulin expression vectors and the SRP9, SRP14, SRP54, SRPRalpha, SRPRbeta, SEC61A1, SEC61B and SEC61G coding sequences were as described by Le Fourn et al. (*Metab. Eng., Epub 2013 Feb 1*).

The GFP, immunoglobulin or secretion proteins were expressed using a eukaryotic expression cassette composed of a human CMV enhancer and human GAPDH promoter upstream of the coding sequence followed by a SV40 polyadenylation signal, the human gastrin terminator and a SV40 enhancer (see Le Fourn et al., 2013). Expression cassettes and/or MAR elements were

inserted between the ITR sequences or in the bacterial vector backbone as illustrated in Figure 1 and in figure legends using standard cloning methods.

Cell culture and transfection analysis

The CHO DG44 cell line was cultivated in DMEM: F12 (Gibco) supplemented with Hypoxanthine/Thymidine (HT, Gibco) and 10% fetal bovine serum (FBS, Gibco). Transfections were performed using PEI (JetPRIME, Polyplus Transfection), according to the manufacturer's instructions. Cells were transfected with various amounts of pDNA sources of PB transposase (ranging from 0 to 1500 ng) for titration experiments or co-transfected with the optimal ratio of 300 ng of PB transposase expression plasmid and 300 ng of transposon donor plasmid. Two days after the transfection, cells were transferred to several Petri dishes depending on the experiment. For analysis of unselected transfected CHO cells, cells were replated without antibiotic selection for 3 weeks and the percentage of fluorescent cells and the fluorescence intensity of GFP positive cells were determined by FACS analysis using a CyAn ADP flow cytometer (Beckman Coulter). For gene copy number analysis of unselected cells, stable GFP positive CHO cells were sorted using a FACSAriaII. For antibiotic resistant colony-counting assays, 50,000 transfected cells were seeded in 100 mm plates and selected with 5 µg/ml puromycin for 2 weeks. Then, resistant colonies were fixed and stained in 70% EtOH 0,7% Methylene Blue for 10 min, and colonies >0.5 mm in diameter were counted. For GFP expression studies, cells were selected for two weeks before GFP fluorescence FACS analysis as described above.

CHO-M cells were maintained in suspension culture in SFM4CHO Hyclone serum-free medium (ThermoScientific) supplemented with L-glutamine (PAA, Austria) and HT supplement (Gibco, Invitrogen life sciences) at 37°C, 5% CO2 in humidified air. Transposon donor plasmids were transferred in these cells by electroporation according to the manufacturer's recommendations (Neon devices, Invitrogen). Quantification of immunoglobulin secretion was performed from batch cultures as described previously (see Le Fourn et al., 2013). Briefly, cell populations expressing immunoglogulins were evaluated in batch cultivation into 50 ml minibioreactor tubes (TPP, Switzerland) at 37°C in 5% CO2 humidified incubator for 7 days. Immunoglobulin concentrations in cell culture supernatants were measured by sandwich ELISA.

qPCR gene copy number assays

Total DNA was isolated from CHO stable cell pools following transposition assays using the DNeasy Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The copy number of genome-integrated transgenes was assessed using 6 ng of genomic DNA by quantitative PCR using the SYBR Green-Taq polymerase kit from Eurogentec Inc and ABI Prism 7700 PCR machine (Applied Biosystems). The GFP-Forward: ACATTATGCCGGACAAAGCC and GFP-Reverse: TTGTTTGGTAATGATCAGCAAGTTG primers were used to quantify the GFP gene, while primers B2M-Forward: ACCACTCTGAAGGAGCCCA and B2M-Reverse: GGAAGCTCTATCTGTGTCAA were used to amplify the Beta-2 microglobulin gene. For the amplicon generated by the B2M primers, one hit was found per CHO haploid genome after alignment to our CHO genome assembly using NCBI BLAST software. As CHO are near-diploid cells, it was estimated that B2M is present at 2 copies per genome. The ratios of the GFP target gene copy number were calculated relative to that of the B2M reference gene, as described previously.

Sorting and assay of immunoglobulin-expressing cells

To magnetically sort IgG-expressing cells, transfected CHO-M cells were seeded at a cell density of $3x10^5$ cells per ml in SFM4CHO medium (Thermo Scientific) supplemented with 8mM L-glutamine and 1x HT supplement (both from Gibco), referred to as Complete Medium. After 4 days in culture, $2x10^6$ cells were washed, re-suspended in PBS and incubated with a biotinylated human IgG (KPL216-1006) at a final concentration of 3 µg/ml, together with 30 µl pre-washed MyOne T1 streptavidin-coated Dynabeads (Invitrogen), on a rotary wheel for 30 minutes at room temperature. The cell and bead mix was then placed on a magnet to separate labeled cells from non-labeled cells. The beads were washed 4 times with a phosphate buffer saline (PBS) solution. After the final PBS wash, the beads and cells were re-suspended in 500 µl pre-warmed Complete Medium, transferred to a 24 well plate and incubated at 37 °C with 5% CO_2. After 24 h the magnetically-sorted polyclonal cells were separated from the beads and incubation was continued until the cells were of a sufficient density for expansion in 50 mL TPP spin tube bioreactors (TECHNO PLASTIC PRODUCTS AG, Switzerland).

Alternatively, two clones were isolated from non-sorted and non-selected populations expressing each of the three IgGs using a ClonePix device. Briefly, semi-solid media was used to immobilize single cells, and colonies secreting high amounts of IgG were picked ten days post-embedding. These cell lines were passaged every 3-4 days in spin tube bioreactors at a density of $3x10^5$ cells/ml in a peptone-containing growth medium (Hyclone SFM4CHO

supplemented with 8 mM glutamine) in a humidified incubator maintained at 37°C and 5% CO2, with orbital shaking at 180 rpm.

IgG titers were determined from cells seeded at a cell density of 1x 10^5 cells per ml and grown for 6 days in 5 ml of Complete Medium in 50 ml Spin tube bioreactors when assessing polyclonal cell populations. Alternatively, shake flask cultures of clonal populations were inoculated at a density of $3x10^5$ cells/ml into SFM4CHO media to initiate the fed batch production process. Fed batch production assays were performed with 25 ml of culture volume in 125ml shake flasks or 5 ml in 50 ml TPP culture tubes in humidified incubators maintained at 37°C and 5% CO2 with shaking at 150 rpm (125 ml shake flask and spin tubes). The production was carried out for ten days by feeding 16%, of the initial culture volume of chemically defined concentrated feed (Hyclone, Cell Boost 5, 52 g/l) on days zero, three and six to eight. No glutamine and glucose feeding was applied during the culture run. The viability and viable cell density (VCD) of the culture was measured daily using a GUAVA machine (Millipore). A double sandwich ELISA assay was used to determine MAb concentrations secreted into the culture media.

Plasmids and relative quantitative PCR analysis

Cloning vectors used in this study are the Selexis mammalian expression vectors SLXplasmid_082. The luciferase sequence of pGL3-Control Vector (Promega) was replaced by a eukaryotic expression cassette composed of a human CMV enhancer and human GAPDH promoter upstream of the EGFP coding sequence followed by a SV40 polyadenylation signal, the human gastrin terminator and a SV40 enhancer. Two human MAR-derived genetic elements are flanking the expression cassette and a puromycin resistance gene expressed from the SV40 promoter, whereas the SLXplasmid_082 differ by the type of the MAR element located upstream of the expression cassette (hMAR 1-68 and hMAR X-29; Girod et al., 2007) .

The trastuzumab and infliximab heavy and light chains cDNAs were cloned in a expression vector to replace EGFP. A vector carrying both the heavy and light chain expression cassette of each IgG was made by combining heavy and light chain expression cassettes together on one plasmid vector. The signal peptide sequence of all heavy and light chains are identical, as are the constant portions of the light chains. The constant portions of the heavy chains differ at several amino acid positions (DEL vs EEM variants).

PCR amplification primers and GenBank accession numbers of the SRP9, SRP14, SRP54, SRPRalpha, SRPRbeta, SEC61A1 and SEC61B cDNAs are listed elsewhere herein. The PCR products encoding secretion proteins were cloned into a vector to replace the EGFP sequence.

When multiple secretion proteins were co-expressed, the inverted terminal sequences of the piggyBac transposon were integrated into vectors to bracket the expression cassette, and the resulting vectors were co-transfected with a piggyBac transposase expression vector to improve transgene integration and obviate the need for antibiotic selection.

A typical PB transposase expression vector is pCS2+U5V5PBU3 which contains the PB transposase coding sequence surrounded by the 5' and 3' untranslated terminal regions (UTR) of the Xenopus laevis β-globin gene was used in related experiments. This plasmid was constructed as follows: the 3' UTR 317 bp fragment from pBSSK/SB10 was inserted into pCS2+U5 (Invitrogen/Life Technologies, Paisley, UK) to yield pCS2+U5U3. The PB transposase coding sequence (2067 bp, GenBank accession number: EF587698) was synthesized by ATG:biosynthetic (Merzhausen, Germany) and cloned in the pCS2+U5U3 backbone between the two UTRs. The PB control vector corresponds to the unmodified pCS2+U5 plasmid.

Different transposons vectors were generated by introducing the PB 235 bp 3' and 310 bp 5' inverted terminal repeats (ITRs), synthesized by ATG:biosynthetic (Merzhausen, Germany), into the pBluescript SK- plasmid (pBSK ITR3'-ITR5'). The neomycin phosphotransferase gene (Neo^R), under the control of the SV40 promoter from pRc/RSV plasmid (Invitrogen/Life Technologies), was then inserted between the two ITRs. The MAR 1-68 and MAR X-29 elements, the puromycin resistance and GFP genes used in this study were as previously described (Girod et al. 2007; Grandjean et al. 2011; Hart and Laemmli 1998). The immunoglobulin expression vectors and the SRP9, SRP14, SRP54, SRPRalpha, SRPRbeta, SEC61A1 and SEC61B coding sequences are described herein. The secretion proteins were expressed using a eukaryotic expression cassette composed of a human CMV enhancer and human GAPDH promoter upstream of the coding sequence followed by a SV40 polyadenylation signal, the human gastrin terminator and a SV40 enhancer. Expression cassettes and/or MAR elements were inserted between the ITR sequences or in the bacterial vector backbone using standard cloning methods.

PiggyBac transposon systems including appropriate 3' and 5' ITRs as well as transposase are, e.g., available from SYSTEM BIOSCIENCE.

For relative quantitative PCR analysis, total RNA was extracted from 1 x 10^5 cells and reverse transcribed into cDNA using the FastLane Cell cDNA Kit (Qiagen) according to the

manufacturer's instructions. The expressions of SRP14 and GAPDH were quantified by qPCR using the Rotor Gene Q (Qiagen) and the LightCycler®480 SYBR Green I Master (Roche) using primers listed in Suppl. Table S1. Messenger RNA levels of SRP14 were normalized to that of GAPDH using the Rotor-Gene Q Series Software (Qiagen).

Cell culture, stable transfection and subcloning of CHO cell lines

Suspension chinese hamster ovary cells (CHO-K1) were maintained in SFM4CHO Hyclone serum-free medium (ThermoScientific) supplemented with L-glutamine (PAA, Austria) and HT supplement (Gibco, Invitrogen life sciences) at 37°C, 5% CO2 in humidified air. CHO-K1 cells were transfected with trastuzumab or infliximab heavy and light chains expression vectors bearing puromycin resistance gene by electroporation according to the manufacturer's recommendations (Neon devices, Invitrogen). Two days later, the cells were transferred in T75 plates in medium containing 10 µg/ml of puromycin and the cells were further cultivated under selection for two weeks. Stable individual cell clones expressing Trastuzumab and Infliximab IgG were then generated by limiting dilution, expanded and analysed for growth performance and IgG production levels. Trastuzumab and Infliximab IgG-producing cell clones expressing the highest IgG levels were selected for further biochemical experiments. Some of these clones were then co-transfected with the SRP14 expressing vector and a plasmid bearing the neomycin resistance gene by electroporation. Cells were then cultivated in medium containing 300 µg/ml of G418 for two weeks as described above. Stable clones were isolated by limited dilution and SRP14 expression was confirmed by Q-PCR assays before culture expansion for biochemical analysis.

Batch and Fed-Batch cultivation

Growth and production performances of individual clones expressing trastuzumab and infliximab were evaluated in batch cultivation into 50-ml minibioreactor (TPP, Switzerland) at 37°C in 5% CO2 humidified incubator for 7 days. At day 3, day 4 and day 7 of the cell cultivation, cell density and viability were determined using the Guava EasyCyte flow cytometry system (Millipore). IgG titer in cell culture supernatants was measured by sandwich ELISA. Cell density ($Cv.ml^{-1}$) and IgG titer values ($\mu g.ml^{-1}$) were plotted at the indicated process time sampling day. The specific IgG productivity of the Trastuzumab and Infliximab expressing clones was determined as the slope of IgG concentration versus integral number of viable cell (IVCD) calculated from day 3 to day 7 (production phase), and expressed as pg per cell and per day (pcd).

For fed-batch production cultures, cells were seeded at 0.3x106 cells/ml into 125 ml shake flasks in 25 ml of SFM4CHO Hyclone serum-free medium. Cultures were maintained at 37°C and 5% CO2 under agitation. Cultures were fed in a daily based with a commercial Hyclone Feed (ThermoScientific). Cell densities and IgG production were daily evaluated.

Proteins Expression and Aggregation Analysis

Soluble cytoplasmic proteins were extracted by permeabilizing cells with 1% Triton X-100 in PBS buffer in presence of a proteases inhibitor cocktail (Roche, inc). After incubation 30 min on ice, cells were centrifuged 10 min at 14,000 rpm. The supernatant was referred to as the "soluble cytosolic and ER proteins" fraction. The pellet was dissolved by sonication in urea Laemmli buffer (62.5 mM Tris, 2% SDS, 8 M Urea, 5 % glycerol, bromophenol blue dye), yielding the aggregated and vesicular insoluble protein fraction. The soluble and insoluble fractions were then adjusted in Laemmli buffer containing or not 2-mercaptoethanol and boiled 8 min at 95°C. Reducing and non-reducing samples were separated on 10% or 4-10% gradient acrylamide gels by sodium dodecyl sulfate polyacrylamide gene electrophoresis (SDS-PAGE), respectively.

Proteins were then blotted onto a nitrocellulose membrane. After blocking in 5 % milk diluted in TBS-Tween (20 mM Tris, 0.5 M NaCl, 0.1% Tween 20), membranes were analysed for different proteins using the following primary antibodies: anti-human IgG (H+L)-HRP conjugated donkey antibody (JK immunoresearch, #709 035 149, 1:5000), anti-human BiP rabbit polyclonal antibody (Cell signaling, BiP, C50B12, 1:2000), anti-human CHOP mouse monoclonal antibody (Cell signaling, CHOP, L63F7, 1:500), anti-human GAPDH goat polyclonal antibody. After overnight incubation at 4°C, each blot was probed with HRP conjugated anti-rabbit IgG or anti-mouse IgG (Cell signaling, 1:20000). Specific proteins recognized by each antibody were detected using ECL reagents and exposure to ECL film (Amersham Biosciences).

Cycloheximide-based Proteins Chase Experiments

Cycloheximide-based chase experiments were carried out onto high (HP) and low (LP) IgG-producers CHO-K1 clones. Equal numbers of cells were plated into 6-wells plates in complete culture medium supplemented with 100 µM of cycloheximide (Sigma). At various time points, cells were harvested and lysed in PBS, 1% Triton X-100. The Tx-soluble and insoluble fractions were then resolved on 4-10% acrylamide non-reducing SDS-PAGE and immunoblotted with anti-human IgG antibody.

Differential Detergent Fractionation Assays

Fractionation of cytosolic from membrane proteins was performed by differential detergent extraction of cell pellet. Cells were first washed in 1 ml PBS, and the plasma membrane of Hp and LP cells was permeabilized in KHM buffer (110mM KAc, 20mM HEPES, 2mM MgCl2, pH 7.2) containing 0,01% digitonin (Sigma) for 10 min in presence or not of 1% of Triton X-100. Semi-permeabilized cells were washed once in KHM buffer and Trypsin was added to 50 µg/ml 10 min at room temperature to digest the soluble proteins. Trypsin digestion was stopped by the addition of 1mM PMSF and 4mM AEBSF. Cells were collected by centrifugation and soluble proteins were extracted in presence of Triton X-100 and protease inhibitors as described in section 2.4. Reducing Laemmli buffer containing 2-mercaptoethanol was added to the pellet and supernatant fractions, which were then subjected to 8% SDS-PAGE. Immunoblotting was performed to detect IgG and BiP proteins.

Cross-linking of proteins and western blotting analysis

Infliximab LP cells were washed once in PBS and incubated in with or without 1mM of the dithiobis(succinimidyl propionate) (DSP) cross-linker (ThermoScientific) for 30 min on ice. Cross-linking was quenched by the addition of 50mM of Tris-HCl (pH 7.4) for 10 min before protein extraction in 1% Triton X-100 containing PBS buffer. After centrifugation 10 min at 14,000 rpm in a microfuge, the Triton X-100 insoluble fraction or whole protein extract were analyzed by SDS-PAGE under reducing condition, immunoblotted and probed with anti-BiP and anti-LC antibodies. Equal amounts of Tx-insoluble fraction proteins were analyzed in parallel.

It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a variety of embodiments, only a few of which are disclosed herein. It will be apparent to the artisan that other embodiments exist and do not depart from the spirit of the invention. Thus, the described embodiments are illustrative and should not be construed as restrictive.

TABLE A: EXEMPLARY LIST OF TEP PROTEINS EXPRESSED USING TRANSPOSON VECTORS

Protein of protein secretion pathway	Name	Origin	Exp or KD[1]	NCBI Reference Sequence[2]	NCBI Reference Sequence[2]	Sequence variation[2]
hSRP14	Homo sapiens Signal Recognition Particle 14kDa (Homologous Alu RNA Binding Protein)	Human	exp	NM_003134.4	NP_003125.3	C to G at CDS position 370 (Pro to Ala at aa position 124) [SEQ ID NO.: 12/13]
hSEC61A1	Homo sapiens Sec61 alpha 1 subunit (S. cerevisiae)	Human	exp	NM_013336.3	NP_037468.1	[SEQ ID NO.: 14/15]
hSEC61B	Homo sapiens Sec61 beta subunit	Human	exp	NM_006808.2	NP_006799.1	[SEQ ID NO.: 16/17]
hSEC61G	Homo sapiens Sec61 gamma subunit	Human	exp	NM_014302.3	NP_055117.1	[SEQ ID NO.: 18/19]
hSRP54	Homo sapiens signal recognition particle 54kDa	Human	exp	NM_003136.3	NP_003127.1	[SEQ ID NO.: 20/21] IN LAST DOC
hSRP9	Homo sapiens signal recognition particle 9kDa	Human	exp	NM_001130440.1	NP_001123912.1	[SEQ ID NO.: 22/23] IN LAST DOC
hSRPRalpha	Homo sapiens Signal Recognition Particle Receptor, A Subunit	Human	exp	NM_003139.3	NP_003130.2	[SEQ ID NO.: 24/25]
hSRPRbeta	Homo sapiens Signal Recognition Particle Receptor, B Subunit	Human	exp	NM_021203.3	NP_067026.3	[SEQ ID NO.: 26/27]
hCANX	Homo sapiens calnexin	Human /CHO	exp/ KD	NM_001746.3	NP_001737.1	[SEQ ID NO.: 28/29]

TABLE A

Proteins of one of the recombination pathways	Name	Origin	Exp or KD[1]	NCBI Reference Sequence[2]	NCBI Reference Sequence[2]	Sequence variation[2]
hRAD51B	Homo sapiens RAD51 paralog B	Human /CHO	exp/ KD	U84138.1	AAC39723.1	
Protein processing and metabolic proteins	**Name**	**Origin**	**Exp or KD[1]**	**NCBI Reference Sequence[2]**	**NCBI Reference Sequence[2]**	**Sequence variation[2]**
hDerlin1	Homo sapiens degradation in endoplasmic reticulum protein 1	Human	exp	NM_024295.4	NP_077271.1	
hHNF-1a	Transcription Factor 1, Hepatic; LF-B1, Hepatic Nuclear Factor (HNF1), Albumin Proximal Factor	Human	exp	NM_000545.5	NP_000536.5	C to G at CDS position 51 (silent substitution), A to C at CDS position 79 (Ile to Leu at aa position 27), C to T at CDS position 1375 (silent substitution), G to A at CDS position 1460 (Ser to Asn at aa position 487)
hHNF-1b var1	Homo sapiens HNF1 homeobox B (HNF1B) variant 1	Human	exp	NM_000458.2	NP_000449.1	
hHNF-1b var2	Homo sapiens HNF1 homeobox B (HNF1B) variant 2	Human	exp	HM116553.1	ADM43490.1	
hHNF-4a	Hepatocyte Nuclear Factor 4, Alpha	Human	exp	NM_000457.3	NP_000448.3	
hVKORC1	Homo sapiens Vitamin K Epoxide Reductase Complex, Subunit 1	Human	exp	NM_024006.4	NP_076869.1	

TABLE A

hUCP2	Homo sapiens uncoupling protein 2 (mitochondrial, proton carrier)	Human	exp	NM_003355.2	NP_003346.2	C to T at CDS position 164 (Ala to Val at aa position 55)
hUCP4	Homo sapiens uncoupling protein 4 (mitochondrial, proton carrier)	Human	exp	NM_004277.3	NP_004268.3	[SEQ ID NO: 30/31]
hCMPSAT	Homo sapiens CMP-sialic acid transporter	Human	exp	NM_006416.4	NP_006407.1	[SEQ ID NO: 32/33]
hLMAN1	Homo sapiens lectin, mannose-binding, 1	Human	exp	NM_005570.3	NP_005561.1	
hMCFD2	Homo sapiens multiple coagulation factor deficiency 2	Human	exp	NM_001171506.2	NP_001164977.1	
CD-rST6Gal1	Rat beta-galactoside alpha2,6-sialyltransferase-derived synthetic sequence	Selexis	exp	N/A	N/A	
rST6Gal1	Rat beta-galactoside alpha2,6-sialyltransferase-derived synthetic sequence	Selexis	exp	N/A	NP_001106815.1	[SEQ ID NO: 34/35]
hCOMSC	Homo sapiens C1GALT1-specific chaperone 1 (C1GALT1C1) variant 2	Human	exp	NM_001011551.2	NP_001011551.1	[SEQ ID NO: 36/37]
hCIRP	Homo sapiens Cold Inducible RNA Binding Protein	Human	exp	NM_001280.2	NP_001271.1	T to C at CDS position 492 (silent substitution)
hGRP78	Homo sapiens Immunoglobulin heavy chain binding protein	Human /CHO	exp/ KD	NM_005347.4	NP_005338.1	
hTMX1	Homo sapiens thioredoxin-related transmembrane protein 1	Human	exp	NM_030755.4	NP_110382.3	T to G at CDS position 492 (silent substitution), G to A at CDS position 648 (silent substitution)

TABLE A

hp97	Homo sapiens VCP valosin containing protein (predicted)	Human	exp	NM_007126.3	NP_009057.1	
hPEPD	Homo sapiens peptidase D (PEPD) variant 1	Human	exp	NM_000285.3	NP_000276.2	C to T at CDS position 1131 (silent substitution
hT-Synthase	Homo sapiens core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase, 1 (C1GALT1)	Human	exp	NM_020156.3	NP_064541.1	A to G at CDS position 107 (Asp to Gly at aa position 36) [SEQ ID NO: 38/39]
h4F2	4F2 heavy chain	Human	exp	AB018010.1	BAA84649.1	
hSPCA1	Homo sapiens ATPase, Ca++ Transporting, Type 2C, Member 1	Human	exp	NM_014382.2	NP_055197.2	
hST6Galnac5	Homo sapiens ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 5	Human	exp	NM_030965.1	NP_112227.1	
hDerlin2	Homo sapiens degradation in endoplasmic reticulum protein 2	Human	exp	NM_016041.3	NP_057125.2	
hMBTPS1	Homo sapiens membrane-bound transcription factor peptidase, site 1	Human	exp	NM_003791.2	NP_003782.1	C to T at CDS position 1317 (silent substitution)
hPRA1	Homo sapiens Rab Acceptor 1 (Prenylated)	Human	exp	NM_006423.2	NP_006414.2	
hTPD52	Homo sapiens tumor protein D52	Human	exp	NM_001025252.1	NP_001020423.1	
hRAB1A var1	Homo sapiens RAB1A (RAB1A) variant 1	Human	exp	NM_004161.4	NP_004152.1	

TABLE A

hRAB1A var2	Homo sapiens RAB1A (RAB1A) variant 2	Human	exp	NM_015543.1	NP_056358.1	
hRAB1B	Homo sapiens RAB1B, member RAS oncogene family	Human	exp	NM_030981.2	NP_112243.1	
hP4HA1	Homo sapiens prolyl 4-hydroxylase, alpha polypeptide	Human	exp	NM_000917.3	NP_000908.2	[SEQ ID NO: 40/41]
hP4HB	Homo sapiens prolyl 4-hydroxylase, beta polypeptide	Human	exp	NM_000918.3	NP_000909.2	[SEQ ID NO: 42/43]
hMGST1	Homo sapiens microsomal glutathione S-transferase 1	Human	exp	NM_145792.1	NP_665735.1	
hCRYAB	Homo sapiens crystallin, alpha B	Human	exp	NM_001885.1	NP_001876.1	
hGILZ	Homo sapiens TSC22 Domain Family, Member 3	Human	exp	NM_198057.2	NP_932174.1	[SEQ ID NO: 44/45]
hCyPB	Homo sapiens Peptidylprolyl Isomerase B (Cyclophilin B)	Human	exp	NM_000942.4	NP_000933.1	[SEQ ID NO: 46/47]
hTIM21	Homo sapiens translocase of inner mitochondrial membrane 21 homolog (TIMM21)	Human	exp	NM_014177.2	NP_054896.2	
hOGC	Homo sapiens Solute Carrier Family 25 (Mitochondrial Carrier; Oxoglutarate Carrier), Member 11	Human	exp	NM_003562.4	NP_003553.2	
hNRF2	nuclear factor, erythroid 2-like 2	Human	exp	NM_006164.3	NP_006155.2	[SEQ ID NO: 48/49]
hHSP47	Homo sapiens Serpin Peptidase Inhibitor,	Selexis	exp	N/A	NP_001226.2	

TABLE A

	Clade H (Heat Shock Protein 47), Member 1, (Collagen Binding Protein 1)					
hLAT1	large neutral amino acid transporter	Human	exp	AB018009.1	BAA84648.1	
hPC	Homo sapiens pyruvate carboxylase	Human	exp	NM_000920.3	NP_000911.2	
hHK1	Homo sapiens hexokinase 1	Human	exp	NM_000188.2	NP_000179.2	G to A at CDS position 1443 (silent substitution) [SEQ ID NO: 50/51]
hPDI	Homo sapiens Protein disulfide isomerase family A, member 2	Human	exp	NM_006849.2	NP_006840.2	G to A at CDS position 1347 (silent substitution) [SEQ ID NO: 52//53]
hBcl-xL	Homo sapiens BCL2-Associated Agonist Of Cell Death	Human	exp	NM_138578.1	NP_612815.1	
hTIM23	Homo sapiens translocase of inner mitochondrial membrane 23 homolog (TIMM23)	Human	exp	NM_006327.2	NP_006318.1	
hPIN1	Homo sapiens peptidylprolyl cis/trans isomerase, NIMA-interacting 1	Human	exp	NM_006221.3	NP_006212.1	[SEQ ID NO: 54/55]
hSEPW1	Homo sapiens selenoprotein W, 1	Human	exp	NM_003009.2	NP_003000.1	[SEQ ID NO: 56/57]
hTOR1A	Homo sapiens torsin family 1, member A (torsin A)	Human	exp	NM_000113.2	NP_000104.1	
hHMGA1	Homo sapiens high mobility group AT-hook 1	Human	exp	NM_145899.2	NP_665906.1	
hP53	Cellular tumor antigen p53 isoform a	Human	exp	NM_000546.5	NP_000537.3	

TABLE A

hNAP1	Homo sapiens nucleosome assembly protein 1-like 1	Human	exp	NM_004537.4	NP_004528.1	A to G at CDS position 471 (silent substitution)
hCOBRA1	Homo sapiens Negative Elongation Factor Complex Member B	Human	exp	NM_015456.3	NP_056271.2	
NLS_DBD_PB	Synthetic DNA-binding protein	Selexis	exp	N/A	N/A	
NLS_DBD_VP 16	Synthetic DNA-binding protein	Selexis	exp	N/A	N/A	]
hATF4	Homo sapiens Activating Transcription Factor 4	Human	exp	NM_001675.2	NP_001666.2	
hCALR	Homo sapiens calreticulin	Human	exp	NM_004343.3	NP_004334.1	[SEQ ID NO: 58/59]
hTAF-Ialpha	Homo sapiens SET translocation (myeloid leukemia-associated) isoform 1	Human	exp	NM_00112282 1.1	NP_0011162 93.1	
hCypA	Homo sapiens Peptidylprolyl Isomerase A (Cyclophilin A)	Human	exp	NM_021130.3	NP_066953.1	
hOct4	Homo sapiens POU Class 5 Homeobox 1	Human	exp	NM_002701.4	NP_002692.2	
hSox2	Homo sapiens SRY (sex determining region Y)-box 2	Human	exp	NM_003106.3	NP_003097.1	
hKlf4	Homo sapiens Kruppel-like factor 4 (gut)	Human	exp	NM_004235.4	NP_004226.3	
hCAV1	Homo sapiens caveolin 1, caveolae protein, 22kDa	Human	exp	NM_001753.4	NP_001744.2	
hCHOP	Homo sapiens CHOP protein (product of DNA-damage-inducible transcript 3)	Human /CHO	exp/ KD	NM_00119505 3.1	NP_0011819 82.1	

TABLE A

DroBiP	Synthetic BIP protein derivative	Selexis	exp	N/A	N/A	[SEQ ID NO: 60] Protein only
hDDOST	Homo sapiens dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit (non-catalytic)	Human	exp	NM_005216.4	NP_005207.2	[SEQ ID NO: 61/62]
hCDX1	Homo sapiens caudal type homeobox 1	Selexis	exp	N/A	NP_001795.2	
hP5CS	Homo sapiens aldehyde dehydrogenase 18 family, member A1 (ALDH18A1)	Human	exp	NM_002860.3	NP_002851.2	
hHSP40	Homo sapiens DnaJ (Hsp40) homolog, subfamily B, Member	Human	exp	NM_006145.1	NP_006136.1	[SEQ ID NO: 63/64]
heIF4A1	Homo sapiens eukaryotic translation initiation factor 4A1	Human	exp	NM_001416.3	NP_001407.1	G to C at CDS position 147 (silent substitution)
hATP5A1	Homo sapiens ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle	Human	exp	NM_001001937.1	NP_001001937.1	[SEQ ID NO: 65/66]
hSERCA2	Homo sapiens ATPase, Ca++ Transporting, Cardiac Muscle, Slow Twitch 2	Human	exp	NM_170665.3	NP_733765.1	[SEQ ID NO: 67/68]
hPHB	Homo sapiens prohibitin	Human	exp	NM_002634.2	NP_002625.1	
hPDIA4	Homo sapiens protein disulfide isomerase family A, member 4	Human	exp	NM_004911.4	NP_004902.1	[SEQ ID NO: 69/70]
hPRPS / MSMB	Homo sapiens beta-microseminoprotein isoform a precursor	Human	exp	NM_002443.3	NP_002434.1	
hH2AFZ	Homo sapiens H2A histone family,	Human	exp	NM_002106.3	NP_002097.1	

TABLE A

	member Z					
hHSC70 /HSPA8	Homo sapiens Heat shock cognate protein 70/ heat shock protein 8	Human	exp	NM_006597.4	NP_006588.1	[SEQ ID NO: 71/72]
hHYOU1	Homo sapiens hypoxia up-regulated 1	Human	exp	NM_006389.3	NP_006380.1	C to T at CDS position 543 (silent substitution), C to T at CDS position 1476 (silent substitution), A to G at CDS position 2235 (silent substitution) [SEQ ID NO: 73/74]
hST3GAL5_var1	Homo sapiens ST3 beta-galactoside alpha-2,3-sialyltransferase 5 isoform 1	Human	exp	NM_003896.3	NP_003887.3	
hST3GAL5_var2	Homo sapiens ST3 beta-galactoside alpha-2,3-sialyltransferase 5 isoform 2	Human	exp	NM_001042437.1	NP_001035902.1	
_NLS_DBD_CDX1_VP16	Synthetic DNA-binding protein	Selexis	exp	N/A	N/A	
hc-Myc	Homo sapiens c-myc proto-oncogene protein	Human	exp	NM_002467.4	NP_002458.2	
hCG40346-ST3Gal1	Homo sapiens ST3 beta-galactoside alpha-2,3-sialyltransferase 1 pseudogene 1 (ST3GAL1P1) on chromosome 4	Human	exp	NG_025114.1	N/A	
hXRCC5	Homo sapiens X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining)	Human	exp	NM_021141.3	NP_066964.1	
hST3GAL1	Homo sapiens ST3 beta-galactoside alpha-2,3-sialyltransferase 1	Human	exp	NM_003033.3	NP_003024.1	G to A at CDS position 819 (silent substitution)

TABLE A

hCMP-SAS	cytidine monophosphate N-acetylneuraminic acid synthetase	Human	exp	NM_018686.4	NP_061156.1	C to T at CDS position 714 (silent substitution) [SEQ ID NO: 75/76]
hERO1-Lalpha	Homo sapiens ERO1-like (ERO1L)	Human	exp	NM_014584.1	NP_055399.1	
hXBP1_var1	Homo sapiens X-box binding protein 1 (XBP1) variant 1	Human	exp	NM_005080.3	NP_005071.2	
hASAH1	Homo sapiens N-acylsphingosine amidohydrolase (acid ceramidase) 1	Human	exp	NM_177924.3	NP_808592.2	T to C at CDS position 737 (Val to Ala at aa position 246)
hSPHK1	Homo sapiens sphingosine kinase 1	Human	exp	NM_021972.3	NP_068807.2	
hGadd45a	Homo sapiens Growth Arrest And DNA Damage-Inducible Protein GADD45 Alpha	Human	exp	NM_001924.3	NP_001915.1	
hASF1A	Homo sapiens anti-silencing function 1A histone chaperone	Human	exp	NM_014034.2	NP_054753.1	
hBeclin-1	Homo sapiens Beclin 1, Autophagy Related	Human	exp	NM_003766.3	NP_003757.1	[SEQ ID NO: 77/78]
hRECQL5_var 3	Homo sapiens RecQ protein-like 5 (RECQL5), transcript variant 3	Human	exp	NM_00100371 6.3	NP_0010037 16.1	
hMLH1 var 1	Homo sapiens mutL homolog 1 (MLH1) variant 1	Human	exp	NM_000249.3	NP_000240.1	T to C at CDS position 1151 (Val to Asp at aa position 384)
hBlimp-1	Homo sapiens PR Domain Containing 1, With ZNF Domain	Human	exp	NM_001198.3	NP_001189.2	
hMGMT	Homo sapiens O-6-methylguanine-DNA methyltransferase	Human	exp	NM_002412.3	NP_002403.2	

TABLE A

hERdj3	Homo sapiens DnaJ (Hsp40) Homolog, *Subfamily B, Member 11*	Human	exp	NM_016306.4	NP_057390.1	[SEQ ID NO: 79/80]
hRECQ1	Homo sapiens RecQ Protein-Like (DNA Helicase Q1-Like)	Human	exp	NM_002907.3	NP_002898.2	
TAT_GyrA	Synthetic DNA-binding protein	Selexis	exp	N/A	N/A	
TAT_GyrB	Synthetic DNA-binding protein	Selexis	exp	N/A	N/A	
TAT_GyrB_EGFP	Synthetic DNA-binding protein	Selexis	exp	N/A	N/A	
NLS_GyrA	Synthetic DNA-binding protein	Selexis	exp	N/A	N/A	
NLS_GyrB	Synthetic DNA-binding protein	Selexis	exp	N/A	N/A	
hAC_Sc	Homo sapiens acid ceramidase isoform b	Selexis	exp	N/A	NP_004306.3	
CHO_AGE	Chinese hamster Ovary N-acylglucosamine 2-epimerase-like	CHO	exp	XM_003497741.1	XP_003497789.1	[SEQ ID NO: 81/82]
hWip1	Wild-Type P53-Induced Phosphatase 1	Human	exp	NM_003620.3	NP_003611.1	[SEQ ID NO: 83/84]
hRTP4	Homo sapiens receptor (chemosensory) transporter protein 4	Human	exp	NM_022147.2	NP_071430.2	C to T at CDS position 192 (silent substitution) [SEQ ID NO: 85/86]
hREEP2	Homo sapiens receptor accessory protein 2	Human	exp	NM_001271803.1	NP_001258732.1	[SEQ ID NO: 87/88]
hDPM1	Homo sapiens dolichyl-phosphate mannosyltransferase polypeptide 1, catalytic subunit	Human	exp	NM_003859.1	NP_003850.1	[SEQ ID NO: 89/90]

TABLE A

hGS	Homo sapiens glutamate-ammonia ligase (glutamine synthetase)	Human	exp	NM_002065.5	NP_002056.2	
hDRiP78	Homo sapiens DnaJ (Hsp40) Homolog, Subfamily C, Member 14	Human	exp	XM_005269201.1	XP_005269258.1	[SEQ ID NO: 91/92]

1. Exp indicates expression of the indicated human protein whereas KD indicates the knock down of the CHO protein. The term Selexis indicates a previously unknown sequence engineered by the applicants.
2. The sequence of the expressed proteins is as indicated by the stated NCBI entry, unless variant sequences were expressed, and the variation are indicated under Sequence variations. N/A indicates a previously unknown sequence engineered by the applicants.

TABLE A

TABLE B: EXEMPLARY LIST OF shRNA EXPRESSED USING, E.G., SPECIFIC PIGGYBAC TRANSPOSON VECTORS

	Vector-RNA placed between ITRs, MAR used	Size (bp)
1	pBSK_ITR_BT+_shRNA1_ccLDHA_X29_ITR claim 1 &25	8984
2	pBSK_ITR_BT+_shRNA2_ccLDHA_X29_ITR	8984
3	pBSK_ITR_BT+_shRNA1_ccMEK1_X29_ITR	8985
4	pBSK_ITR_BT+_shRNA3_ccMEK1_X29_ITR	8985
5	pBSK_ITR_BT+_shRNA1_ccMEK2_X29_ITR	8987
6	pBSK_ITR_BT+_shRNA3_ccMEK2_X29_ITR	8987
7	pBSK_ITR_BT+_shRNA2_ccDNMT1_X29_ITR	8987
8	pBSK_ITR_BT+_shRNA3_ccDNMT1_X29_ITR	8987
9	pBSK_ITR_BT+_shRNA4_ccDNMT1_X29_ITR	8987
10	pBSK_ITR_BT+_shRNA1_ccDNMT3a_X29_ITR	8987
11	pBSK_ITR_BT+_shRNA2_ccDNMT3a_X29_ITR	8987
12	pBSK_ITR_BT+_shRNA3_ccDNMT3a_X29_ITR	8987
13	pBSK_ITR_BT+_shRNA1_ccRad51_X29_ITR	8983
14	pBSK_ITR_BT+_shRNA2_ccRad51_X29_ITR	8983
15	pBSK_ITR_BT+_shRNA3_ccRad51_X29_ITR	8983
16	pBSK_ITR_BT+_shRNA1_ccIFNAR1_X29_ITR	8987
17	pBSK_ITR_BT+_shRNA2_ccIFNAR1_X29_ITR	8987
18	pBSK_ITR_BT+_shRNA3_ccIFNAR1_X29_ITR	8983
19	pBSK_ITR_BT+_shRNA1_ccP2X7_X29_ITR	8983
20	pBSK_ITR_BT+_shRNA2_ccP2X7_X29_ITR	8983
21	pBSK_ITR_BT+_shRNA3_ccP2X7_X29_ITR	8983
22	pBSK_ITR_BT+_shRNA1_mPEPCK_X29_ITR	8987
23	pBSK_ITR_BT+_shRNA2_mPEPCK_X29_ITR	8987
24	pBSK_ITR_BT+_shRNA3_mPEPCK_X29_ITR	8987
25	pBSK_ITR_BT+_shRNA1_ccBiP_X29_ITR	8983
26	pBSK_ITR_BT+_shRNA2_ccBiP_X29_ITR	8983
27	pBSK_ITR_BT+_shRNA3_ccBiP_X29_ITR	8983
28	pBSK_ITR_BT+_shRNA1_ccP53_X29_ITR	8987
29	pBSK_ITR_BT+_shRNA2_ccP53_X29_ITR	8987
30	pBSK_ITR_BT+_shRNA3_ccP53_X29_ITR	8987
31	pBSK_ITR_BT+_shRNA1_pre-miR-466h-5p_X29_ITR	8985
32	pBSK_ITR_BT+_shRNA1_miR-466h-5p_X29_ITR	8985

TABLE C: LIST OF EXAMPLES OF siRNAs (SENSE STRAND) AND EXAMPLES OF shRNAS CREATED FROM CORRESPONDING siRNAS

53BP1_1 UCAGAAUGAUGACAAAGUA
53BP1_2 GAGCAAGGAGACAAUAAUA
53BP1_3 CAAAGACAUCCCUGUUACA

Brca1_1 CCACGUAACUGAAAUUAUA
Brca1_2 AAGGCUGAGUUCUAUAAUA
Brca1_3 AGAGCCAAAUGAACAAAGA

Brca2_1 GAAGCUGUUUACAGAAUGA
Brca2_2 CAAUGACUAUACAGACAAA
Brca2_3 AACAGACGGUUGCCAUAAA

cycD1_1 UGGAACUCCUUCUGGUGAA
cycD1_2 CGCACUUUCUUUCCAGAGU
cycD1_3 UGCCAGAGGCGGAUGAGAA

DNA-PKcs_1 GGAUCGAGCUGUUCAGAAA
DNA-PKcs_2 AGAUGAUGUUCACUCUAAA
DNA-PKcs_3 AUCCAUCGGUAUCUUUAAA

Ku70_1 GGUGCCCUUUACUGAGAAA
Ku70_2 AAAGCCCAAGGUAGAGUUA
Ku70_3 ACAUUUCCAAGACACAAUU

Ku80_1 GAAACUGUCUAUUGCUUAA
Ku80_2 CCAUAGGGAAGAAGUUUGA
Ku80_3 GGAUUCCUAUGAGUGUUUA

LigIV_1 AGAGCCUCCUUCAGUUAAU
LigIV_2 CUAUACAGCAGGUAAAUGA
LigIV_3 AGAGGUAUGAUAUCCUUAA

Rad51_1 GUGCCAAUGAUGUGAAGAA

*Corresponding Rad51 shRNA coding sequence:
ACAAGCTTGTGCCAATGATGTGAAGAATTCAAGAGATTCTTCACATCATTGGCACTCTAGAG
TCGGGGCGGCCGGCC

Rad51_2 GGGAAUUAGUGAAGCCAAA

**Corresponding Rad51 shRNA coding sequence :
ACAAGCTTGGGAATTAGTGAAGCCAAATTCAAGAGATTTGGCTTCACTAATTCCCTCTAGAG
TCGGGGCGGCCGGCC

TABLE C

Rad51_3 GGCGUUCAGAAAUCAUACA

**Corresponding Rad51 shRNA coding sequence :
ACAAGCTTGGCGTTCAGAAATCATACATTCAAGAGATGTATGATTTCTGAACGCCTCTAGAG
TCGGGGCGGCCGGCC

Rad51b_1 ACAGCCUAUGAUAUAAAGA
Rad51b_2 CAAGUUCUUGGCCAAACAA
Rad51b_3 GUACCUGGCUGAGGAAUUU

Rad51c_1 UGAUCAGCCUGGCAAAUAA
Rad51c_2 AGAGGAAGCUUUAGAAACU
Rad51c_3 GGAUGAAGAACACCAGAAA

Rad51d_1 ACGGAGCAGACCUAUAUGA
Rad51d_2 CCCAAGAUGAGGAGAAACA
Rad51d_3 GCCUGGACAAACUACUUGA

Rad52_1 UGAGAUGUUUGGUUACAAU
Rad52_2 ACUGCAUUCUGGACAAAGA
Rad52_3 CCCUGAAGACAACCUUGAA

Rad54_1 AGAAGACCUGCUAUAUUUA
Rad54_2 CAUCAGAUAUCCUCUCUAA
Rad54_3 GAAGCUAUGUAACCAUCCA

Xrcc2_1 GAAGUGUUCUCAGCUCCUA
Xrcc2_2 CAACACAAAGUCUAAUGCA
Xrcc2_3 AUCAGAGGGUGGACUGCAA

Xrcc3_1 CCACAUCUUCAUCGAGCAU
Xrcc3_2 ACGGUGGAGGAGCAAGAGU
Xrcc3_3 GAUCAGAUUCAGCAACCAC

Xrcc4_1 AUAUGCUGAUGAAUUGAGA
Xrcc4_2 CUGAAAGAUGUCUCAUUUA
Xrcc4_3 AUGAGCACCUGCAGAAAGA

Neg. Control 1 AGGUAGUGUAAUCGCCUUG
Neg. Control 2 GACGACUCACAUACGUAAA
Neg. Control 3 GAAUAUAUCGCGAAAUGUA

(*) (**) (***) illustrate the structure of a shRNA-encoding DNA sequence from the siRNA sequences listed using the Rad51-targeted molecules as examples. The DNA sequence corresponding to the shown siRNA strand is underlined, whereas the complementary sequence allowing hairpin formation is underlined twice.

TABLE C

TABLE D: SELECTED GENES OF CERTAIN RECOMBINATION PATHWAYS

Recombination pathway	Target gene [sequence identifiers in brackets identify sequences of example genes]	Full name of target gene	Transfected vector		N° exp.
			GFP[1]	MAR-GFP[1]	
NHEJ	Xrcc4	X-ray repair complementing defective repair in Chinese hamster cells 4	0	0	3
	Ku70	X-Ray Repair Complementing Defective Repair In Chinese Hamster Cells 6	-	0	3
	Ku80	X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining)	-	0	3
	LigIV	Ligase IV, DNA, ATP-Dependent	0	0	3
	DNA-PKcs	Protein Kinase, DNA-Activated, Catalytic Polypeptide	-	0	3
	53BP1 **[SEQ ID NO.:109]**	Tumor suppressor p53-binding protein 1	+	+	3
HR	Rad51 **[SEQ ID NO.:93]**	RAD51 recombinase // DNA repair protein RAD51	+++	+++	14
	Rad51B **[SEQ ID NO.:94]**	RAD51 paralog B // DNA repair protein RAD51 homolog 2	-	0	3
	Rad51C **[SEQ ID NO.:95]**	RAD51 paralog C // DNA repair protein RAD51 homolog 3	+	+	4

TABLE D

HR *(continued)*	Rad51D **[SEQ ID NO.:96]**	RAD51 paralog D // DNA repair protein RAD51 homolog 4	0	+++	4
	Rad52 **[SEQ ID NO.:97]**	RAD52 // DNA repair protein RAD52	+	++	3
	Rad54 **[SEQ ID NO.:98]**	RAD54 // DNA repair and recombination protein RAD54	+	++	3
	Xrcc2 **[SEQ ID NO.:99]**	X-ray repair complementing defective repair in Chinese hamster cells 2 // DNA repair protein XRCC2	0	+	3
	Xrcc3 **[SEQ ID NO.:100]**	X-ray repair complementing defective repair in Chinese hamster cells 3 // DNA repair protein XRCC3	+	+++	3
	Brca1 **[SEQ ID NO.:101]**	breast cancer 1, early onset // breast cancer type 1 susceptibility protein	+	++	3
	Brca2 **[SEQ ID NO.:102]**	breast cancer 2, early onset // breast cancer type 2 susceptibility protein	--	0	3
	Cyclin D1 **[SEQ ID NO.:103]**	Cyclin D1	0	+++	3
	Bard1 **[SEQ ID NO.:106]**	BRCA1 associated RING domain 1 // BRCA1 associated RING domain 1	+	++	2
MRN	Mre11 **[SEQ ID NO.:108]**	Mre11 = meiotic recombination 11 // Double-strand break repair protein MRE11	0	+++	6

TABLE D

MRN ***(continued)***	Rad50	RAD50 Homolog (S. Cerevisiae)	--	+	3
	Nbs1	Nibrin	---	+	3
MMEJ	Ercc1 **[SEQ ID NO.:104]**	excision repair cross-complementing rodent repair deficiency, complementation group 1 // DNA excision repair protein ERCC-1	+	+	3
	Xpf	excision repair cross-complementing rodent repair deficiency, complementation group 4	0	0	1
	Pol theta	Polymerase (DNA directed), theta	0	0	3
	Ligase I **[SEQ ID NO.:107]**	DNA ligase 1 // DNA ligase 1	+	+	1
	Ligase III	Ligase III, DNA, ATP-Dependent	0	0	5
	Xrcc1	X-ray repair complementing defective repair in Chinese hamster cells 1	0	0	3
	CtIP	mediator of DNA-damage checkpoint 1 // Mediator of DNA damage checkpoint protein 1	+	+	1
	PARP1	mediator of DNA-damage checkpoint 1 // Mediator of DNA damage checkpoint protein 1	+	+	3
	POLD3		0	0	2

TABLE D

DNA repair proteins	MDC1 **[SEQ ID NO.:105]**	mediator of DNA-damage checkpoint 1 // Mediator of DNA damage checkpoint protein 1	+	-	1
	MSH2	mutS homolog 2	0	0	3

[1] Type of effect of the knock-down on GFP transgene expression; + positive effect, ++ statistically significant positive effect ($p<0.05$), +++ statistically highly significant positive effect ($p<0.01$), - negative effect, -- statistically significant negative effect ($p<0.05$), --- statistically highly significant negative effect ($p<0.01$).

TABLE D

CLAIMS

1. A recombinant nucleic acid molecule comprising:

(a) a 5' and a 3' transposon-specific inverted terminal repeat (ITR),

(b) at least one nucleic acid sequence encoding a transgene expression processing (TEP) protein or a TEP functional RNA, located between the 5' and 3' ITRs and which is under the control of a promoter, and

(c) at least one MAR (matrix attachment region) element;

in which the molecule does not comprise a selection gene which, in use, is used during expression of a gene of interest.

2. A recombinant nucleic acid molecule according to Claim 1, in which there is provided at least one transgene also located between the 5' and 3' ITRs and which is under the control of a transgene promoter.

3. A recombinant nucleic acid molecule according to Claim 1 or 2, in which the MAR element is located between the 5' and a 3' ITRs.

4. A recombinant nucleic acid molecule according to any preceding claim, in which the TEP protein or TEP functional RNA is a protein or a functional RNA that is, directly or indirectly, involved in integration of nucleic acid sequences into a genome, processing or translation of the transgene RNA product or is involved in ER translocation, secretion, processing, folding, ER-Golgi-plasma membrane transport, glycosylation and/or another post-translational modification of proteins such as transgene expression products.

5. A recombinant nucleic acid molecule according to any preceding claim, in which the TEP protein is a protein of the protein secretion pathway, a protein of the recombination pathway, a protein processing or metabolic protein including chaperones such as BiP, or a combination thereof.

6. A recombinant nucleic acid molecule according to any preceding claim, in which the TEP protein is one or more of the following proteins of the protein secretion pathway:

(a) one or more of the following proteins of the protein secretion pathway: hSRP14, hSec61α1, hSec61β, hSec61γ, hSRP54, hSRP9, hSRPRα, hSRPβ, and hCANX; or

(b) one or more of the following amino acid sequences of proteins of the protein secretion pathway: hSRP14 having SEQ ID NO: 13, hSec61α1 having SEQ ID NO: 15, hSec61β having SEQ ID NO: 17, hSec61γ having SEQ ID NO: 19, hSRP54 having SEQ ID NO: 21, hSRP9 having SEQ ID NO: 23, hSRPRα having SEQ ID NO: 25, hSRPβ having SEQ ID NO: 27, and hCANX having SEQ ID NO: 29 and amino acid sequences having more than 80%, 90%, 95% or 98% sequence identity with the specified sequences.

7. A recombinant nucleic acid molecule according to any preceding claim, in which the TEP protein is one or more of the following protein processing or metabolic proteins: hUCP4, hCMPSAT, rST6Gal1, hCOMSC, hT-Synthase, hP4HA1, hP4HB, hGILZ, hCyPB, hNRF2, hHK1, hPDI, hPIN1, hSEPW1, hCALR, hDDOST, hHSP40, hATP5A1, hSERCA2, hPDIA4, hHSC70 /HSPA8, hHYOU1, hCMP-SAS, hBeclin-1, hERdj3, CHO-AGE, hWip1, hRTP4, hREEP2, hDPM1 and hDRiP78.

8. A recombinant nucleic acid molecule according to any preceding claim, in which the TEP protein is a synthetic BIP protein derivative (DroBiP) having 80%, 90%, 95% or 100% sequence identity with the SEQ ID NO: 60.

9. A recombinant nucleic acid molecule according to any preceding claim, in which the MAR element is selected from SEQ ID NOs: 1 (MAR 1-68), 2 (MAR 1_6), 3 (MAR X_S29), 4 (MAR S4), 5 (chicken lysozyme MAR), is a engineered, in particular rearranged counterpart and/or has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with any of SEQ ID NOs: 1 to 5.

10. A recombinant nucleic acid molecule according to any preceding claim, in which the TEP functional RNA within the cell comprises/consists of nucleic acid sequences encoding a functional RNA, preferably a miRNA or a shRNA, that interferes with the expression of at least one protein of a recombination pathway, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2, Cyclin D1, Ercc, MDC1, Bard1, Ligase 1, Mre 11 and/or 53BP1.

11. A recombinant nucleic acid molecule according to any preceding claim, in which the nucleic acid molecule is at least 5000, 6000, 7000, 8000, 90000 or 10000 bps long.

12. A recombinant nucleic acid molecule according to any preceding claim, in which the 5' and a 3' ITRs are 5' and 3' ITRs of the Sleeping Beauty or preferably PiggyBac Transposon.

13. A recombinant nucleic acid molecule according to any preceding claim, in which the recombinant nucleic acid molecule is part of a vector.

14. A recombinant nucleic acid molecule according to Claim 13, in which the vector comprises a singular MAR element, two or more MAR elements, in which the element(s) are located between the 5' and 3' ITRs.

15. A recombinant nucleic acid according to Claim 13, in which the vector comprises two MAR elements, in which a first MAR element is positioned upstream of the TEP or TEP functional RNA and a second MAR element is positioned downstream of the TEP or TEP functional RNA, in which the first MAR element comprises a MAR 1_6 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO: 2, in particular a rearranged MARs based on MAR 1-6, more in particular elements that have at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO: 8 (MARs 1_6R2) and the second MAR element comprises a MAR 1-68 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO: 1.

16. A recombinant nucleic acid molecule according to Claim 13, in which the vector comprises the singular MAR element, in which the singular MAR element is positioned downstream of the TEP or TEP functional RNA, in which the singular MAR element is a MAR 1-68 or a MAR X-29 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NOs. 1 or 3, in particular a rearranged MAR based on MAR 1-68 or a MAR X-29, in particular an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NOs: 6, 7 or 10 (MARs 1_68R, 1_68R2 or X_29R3) or 9, and is preferably a MAR X-29 element and/or an element that has at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO: 3.

17. A recombinant nucleic acid molecule according to any of Claims 13 to 16, in which the TEP or TEP functional RNA is under the control of an EF1 alpha promoter and is optionally followed by a BGH polyA signal.

18. A recombinant nucleic acid molecule according to any of Claims 13 to 17, in which the vector comprises promoter(s) and/or enhancer(s) or fusions thereof such as GAPDH, CGAPDH, SV40p, CMVp, CHO Ef1 alpha, CHO Actb and/or CHO Hspa5.

19. An *in vitro* method for expressing a TEP or TEP functional RNA comprising providing a recombinant mammalian cell comprising a transgene, and the vector is a vector according to any of Claims 13 to 18 which expresses the TEP or TEP functional RNA, in which the TEP or TEP functional RNA expressed via the vector optionally increases an expression of a transgene in the mammalian cell by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70%.

20. An *in vitro* method according to Claim 19, in which the vector comprises a singular MAR X-29 element and/or a nucleic acid sequence having at least 80%, 90%, 95%, 98%, 99% or 100% sequence identity with SEQ ID NO: 3 and in which, after more than 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 weeks of cultivation, the TEP or TEP functional RNA expressed via the vector increases an expression of the gene of interest by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70%.

21. A recombinant mammalian cell comprising not more than 20, 15, 10 or 5 copies of a recombinant nucleic acid molecule according to any of Claims 1 to 18.

22. A recombinant mammalian cell according to Claim 21 in which recombinant nucleic acid molecule is integrated into the genome of the cell as single copies.

23. A recombinant mammalian cell according to Claim 21 or 22, in which the nucleic acid sequence(s) encoding a TEP protein within the cell comprise(s)/consist(s) of nucleic acid sequences encoding one or more of the proteins according to any of Claims 7, 8 and/or 9.

24. A recombinant mammalian cell according to Claim 23, in which the nucleic acid sequence(s) encoding the TEP functional RNA within the cell comprise(s)/consist(s) of nucleic acid sequence(s) encoding a functional RNA, preferably a miRNA or a shRNA, that

interferes with the expression of at least one recombination protein, preferably an HR gene, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2, Cyclin D1, Ercc1, MDC1, Bard1, Ligase 1, Mre 11 and/or 53BP1.

25. A recombinant mammalian cell according to any of Claims 21 to 24, in which the cell is a primary stem cell, a hamster, a CHO (Chinese hamster ovary) cell, a human cell, or HEK293 cell.

26. An *in vitro* method for transfecting mammalian cells, in particular hamster cells, comprising transfecting the mammalian cells with

(i) at least one, at least two, at least three or at least four of the recombinant nucleic acids molecules according to any of Claims 1 to 12 and/or

(ii) at least one, at least two, at least three or at least four isolated TEP functional RNA and at least one transgene.

27. An *in vitro* method according to Claim 26, in which the transgene is part of a recombinant nucleic acid molecule.

28. An *in vitro* method according to Claim 26 or 27, in which the transgene can be transfected in a second, subsequent transfection.

29. An *in vitro* method according to Claim 28, in which the second, subsequent transfection can also include transfection together with an isolated nucleic acid or mRNA expressing a transposase which recognizes the 5' and the 3' ITR.

30. An *in vitro* method according to any of Claims 26 to 29, in which the cell is transfected with more than one, including at least two, at least three or at least four of the recombinant nucleic acid molecules encoding one, two or three of the TEP proteins specified in Claim 6.

31. An *in vitro* method according to any of Claims 26 to 30, in which each of the nucleic acid molecules is part of a vector and the vectors are co-transfected.

32. An *in vitro* method according to Claim 31, in which the co-transfection, preferably a co-transfection of recombinant nucleic acid molecules encoding proteins SRP14, SRP9 and

SRP54 increases transgene integration and/or expression in the cell relative to a cell not subject to such co-transfection.

33. An *in vitro* method according to any of Claims 26 to 32, in which a number of the mammalian cells that stably express the TEP protein or TEP functional RNA is obtained to obtain recombinant mammalian cells and in which the number of recombinant mammalian cells is independent from the presence of the MAR element.

34. An *in vitro* method according to any of Claims 26 to 33, in which the mammalian cells are transfected a second time.

35. An *in vitro* method according to any of Claims 26 to 34, in which at least 30%, 40% or 45% of the mammalian cells become recombinant mammalian cells and express the transgene.

36. An *in vitro* method according to any of Claims 26 to 35, in which the mammalian cell is transfected with the at least one isolated TEP functional RNA and a vector comprising the at least one transgene.

37. An *in vitro* method according to Claim 36, in which a 3' ITR and 5' ITR flanks the at least one transgene.

38. An *in vitro* method according to Claim 37, in which there is also provided an isolated nucleic acid or mRNA expressing a transposase which recognizes the 5' and the 3' ITR.

39. An *in vitro* method according to any of Claims 26 to 38, in which the transgene is a therapeutic protein such as an immunoglobulin, hormone, cytokine or growth factor.

40. An *in vitro* method according to Claim 26, in which the TEP functional RNA is encoded by a recombinant nucleic acid sequences encoding a shRNA or comprises/consists of a siRNA that interferes with the expression of at least one HR gene, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2, Cyclin D1, Ercc1, MDC1, Bard1, Ligase 1, Mre 11 and/or 53BP1.

41. An *in vitro* method according to any of Claims 26 to 42, in which transgene integration and/or expression is increased in such a cell relative to a cell not transfected with the isolated nucleic acid molecules and/or the at least one of the isolated TEP functional RNAs.

42. A kit comprising in one container at least one vector comprising a recombinant nucleic acid molecule according to any of Claims 1 to 12.

43. A kit according to Claim 42 comprising a second container comprising a vector encoding a compatible transposase and in a further container instruction of how to use the vector or vectors.

44. A kit according to Claim 42, in which more than one vector is provided in one or more containers and in which the TEP proteins are at least two of the following: a chaperone, SRP14, SRP9, or SRP54.

45. A kit according to Claim 42, in which the TEP functional RNA(s) within the vector(s) comprise(s)/consist(s) of nucleic acid sequences encoding a miRNA, siRNA or a shRNA that interferes with the expression of at least one HR gene, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2 Cyclin D1, Ercc1, MDC1, Bard1, Ligase 1, Mre 11 and/or 53BP1, and optionally in a further container siRNA(s) that interferes with the expression of at least one HR gene, such as, but not limited to, Rad51, Rad51B, Rad51C, Rad51D, Xrcc2, Xrcc3, Rad52, Rad54, Brca1, Brca2, Cyclin D1, Ercc1, MDC1, Bard1, Ligase 1, Mre 11 and/or 53BP1.

46. *In vitro* use of a recombinant nucleic acid according to any of Claims 1 to 12 and/or a recombinant mammalian cell according to any of Claims 21 to 24 for increasing transgene integration and/or expression.