Integration sites in CHO cells

TECHNICAL FIELD

[001] The present invention relates to the identification of a genomic integration site for heterologous polynucleotides in Chinese Hamster Ovary (CHO) cells resulting in high RNA and/or protein production. More specifically it relates to CHO cells comprising at least one heterologous polynucleotide stably integrated into the S100A gene cluster of the CHO genome and to methods for the production of said CHO cells. Further, the invention relates to a method for the production of a protein of interest using said CHO cell and to the use of said CHO cell for producing a protein of interest at high yield. Integration within these specific target regions leads to reliable, stable and high yielding production of an RNA and/or protein of interest, encoded by the heterologous polynucleotide.

TECHNOLOGICAL BACKGROUND

[002] Chinese hamster ovary (CHO) cells are the most popular host cells for the recombinant production of therapeutic proteins. Classical cell line development procedures rely on the random integration (RI) of expression vectors followed by selection and screening of subclones for optimal productivity behavior. Random integration is associated with a large heterogeneity in the resulting cell population, owing to unpredictable chromosomal positioning effects, variable copy numbers and stability issues. High producer cells account for only a small proportion of the randomly transfected cells and tend to be outgrown by low producer cells. Hence, a large number of clones need to be screened in order to identify and isolate one individual clone suitable for sustained biopharmaceutical protein production and fermentation process development.

[003] Positional effects on the expression of heterologous genes can result from, e.g., chromatin structure, genomic imprinting or the presence of transcriptional regulator elements, such as genomic enhancer elements, silencer elements or promoter elements in the vicinity of the integration site (C. Wilson *et al*. Annu. Rev. Cell Biol. 1990, 6, 679-714). Many of these elements within the genome are not known or characterized, and the potential of a genomic locus in a cell line development process therefore hard to predict.

[004] By replacing classical random integration with targeted integration (TI) of the protein expression vector into one or more pre-determined genomic locus/loci, these disadvantages can be overcome. Targeted integration makes the cell line development process much more predictable as all subclones will have identical genomic set ups negating the need for extensive screening procedures.

[005] The challenge for a cell line development process that relies on targeted integration lies in the identification of a suitable genomic locus, often called a “hot spot”. The ideal site(s) will support

expression levels without excessive down-regulation, be amplifiable using metabolic selection markers such as DHFR or GS in conjunction with MTX or MSX, and will be located so that integration of transgenes does not negatively impact cell growth or protein product profiles.

[006] The S100A6 gene is part of the S100A gene cluster encoding a group of known calcium-binding proteins, e.g. S100A1, S100A13, S100A14, S100A16, S100A3, S100A2, S100A4, S100A5 and S100A6. The cluster comprises a “side cluster” including the S100A1, S100A13, S100A14 and S100A16 genes and a “main cluster”, which includes the S100A3, S100A4, S100A5 and S100A6 genes.

[007] In the present invention, it is shown that the stable integration of heterologous polynucleotides within the S100A gene cluster of the CHO cell genome increases the production of a heterologous gene product. Specifically, stable integration within the upstream and downstream regions flanking the S100A3/A4/A5/A6 main gene cluster, enables a predictable, high level and stable production of a heterologous gene product, including recombinant proteins, such as antibodies and fusion proteins, or regulatory RNAs, such as shRNAs or miRNAs.

SUMMARY

[008] In an aspect, the present disclosure refers to a Chinese hamster ovary (CHO) cell, comprising at least one heterologous polynucleotide, stably integrated into the S100A gene cluster of the CHO cell genome is described, wherein the at least one heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or the at least one heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2. Preferably the upstream genomic target region corresponds to nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1; and/or the downstream genomic target region corresponds to nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2.

[009] More preferably the upstream genomic target region corresponds to nucleotides 11,820 to 18,720 of SEQ ID NO: 1, nucleotides 13,560 to 18,720 of SEQ ID NO: 1, nucleotides 15,360 to 18,720 of SEQ ID NO: 1 or nucleotides 17,100 to 18,720 of SEQ ID NO: 1; and/or the downstream genomic target region corresponds to nucleotides 660 to 10,260 of SEQ ID NO: 2, nucleotides 1,320 to 10,260 of SEQ ID NO: 2 or nucleotides 1,480 to 10,260 of SEQ ID NO: 2. Even more preferably the upstream genomic target region corresponds to nucleotides 11,820 to 18,380 of SEQ ID NO: 1, nucleotides 13,560 to 18,380 of SEQ ID NO: 1, nucleotides 15,360 to 18,380 of SEQ ID NO: 1 or nucleotides 17,100 to 18,380 of SEQ ID NO: 1; and/or the downstream genomic target region

corresponds to nucleotides 3,180 to 10,260 of SEQ ID NO: 2, nucleotides 4,920 to 9,000 of SEQ ID NO: 2 or nucleotides 6,720 to 8,460 of SEQ ID NO: 2.

[010] In one example, the at least one heterologous polynucleotide is stably integrated into the CHO cell genome as part of an expression cassette. The at least one heterologous polynucleotide may code for a RNA and/or a protein. The RNA may be an mRNA, a miRNA or a shRNA. The protein may be a therapeutic protein, preferably a therapeutic protein selected from the group consisting of an antibody, a fusion protein, a cytokine and a growth factor.

[011] The at least one heterologous polynucleotide may also be a marker gene selected from the group consisting of a reporter gene and a selection marker gene. Preferably the marker gene is stably integrated into the CHO cell genome as part of an expression cassette and the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme such as a site specific nuclease.

[012] The CHO cell according to the invention may be a CHO-DG44 cell, a CHO-K1 cell, a CHO-DXB11 cell, a CHO-S cell, a CHO glutamine synthetase (GS)-deficient cell or a derivative of any of these cells.

[013] In one example, the genomic target region consists of any one of the sequences defined in SEQ ID NO: 1 and/or SEQ ID NO: 2 above or a sequence having at least 80% sequence identity thereto.

[014] The at least one heterologous polynucleotide may be stably integrated into one or both alleles of the S100A gene cluster of the CHO cell genome.

[015] In another aspect, the present disclosure refers to a method for the production of a CHO cell, comprising the steps of (a) providing a CHO cell; (b) introducing a heterologous polynucleotide into said CHO cell, wherein the heterologous polynucleotide is stably integrated into the S100A gene cluster of the CHO cell genome, wherein said heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or said heterologous polynucleotide is integrating downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2.

[016] Preferably the upstream genomic target region corresponds to nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1; and/or the downstream genomic target region corresponds to nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2. More preferably the upstream genomic target region corresponds to nucleotides 11,820 to 18,720 of SEQ ID NO: 1, nucleotides 13,560 to 18,720 of SEQ ID NO: 1, nucleotides 15,360 to 18,720 of SEQ ID NO: 1 or nucleotides 17,100 to 18,720 of SEQ ID NO: 1; and/or the downstream genomic target region corresponds to nucleotides 660 to 10,260 of SEQ ID NO: 2, nucleotides 1,320 to 10,260 of

SEQ ID NO: 2 or nucleotides 1,480 to 10,260 of SEQ ID NO: 2. Even more preferably the upstream genomic target region corresponds to nucleotides 11,820 to 18,380 of SEQ ID NO: 1, nucleotides 13,560 to 18,380 of SEQ ID NO: 1, nucleotides 15,360 to 18,380 of SEQ ID NO: 1, nucleotides 17,100 to 18,380 of SEQ ID NO: 1; and/or the downstream genomic target region corresponds to nucleotides 3,180 to 10,260 of SEQ ID NO: 2, nucleotides 4,920 to 9,000 of SEQ ID NO: 2 or nucleotides 6,720 to 8,460 of SEQ ID NO: 2.

[017] In one example, the genomic target region consists of any one of the sequences defined in SEQ ID NO: 1 and/or SEQ ID NO: 2 above or a sequence having at least 80% sequence identity thereto.

[018] In another example, the at least one heterologous polynucleotide is stably integrated into the CHO cell genome as part of an expression cassette and the expression cassette may be flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme (e.g., a site specific nuclease)

[019] In one example, the at least one heterologous polynucleotide is stably integrated into the CHO cell genome as part of an expression cassette. The at least one heterologous polynucleotide may code for a RNA and/or a protein. The RNA may be an mRNA, a miRNA or a shRNA. The protein may be a therapeutic protein, preferably a therapeutic protein selected from the group consisting of an antibody, a fusion protein, a cytokine and a growth factor.

[020] The at least one heterologous polynucleotide may also be a marker gene selected from the group consisting of a reporter gene and a selection marker gene. Preferably the marker gene is stably integrated into the CHO cell genome as part of an expression cassette and the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme (e.g., a site specific nuclease).

[021] The at least one heterologous polynucleotide may be stably integrated into one or both alleles of the S100A gene cluster of the CHO cell genome.

[022] In one example, the heterologous polynucleotide is introduced into the CHO cell genome using (a) a sequence specific DNA editing enzyme, preferably a site specific nuclease, more preferably selected from the group consisting of zinc finger nucleases (ZFNs), meganucleases, transcription activator-like effector nucleases (TALENs) and CRISPR associated nucleases; or (b) a site-specific recombinase, preferably selected from the group consisting of lambda integrase, PhiC31 integrase, Cre, Dre and Flp.

[023] In another example, the method may further comprise the steps of (a) providing a CHO cell; (aa) introducing a first heterologous polynucleotide into said CHO cell, wherein the first heterologous polynucleotide is a marker gene and is stably integrated into the S100A gene cluster of the CHO cell genome as part of an expression cassette flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme (e.g., a site specific nuclease), wherein (i) said heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or (ii) said

heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2; and introducing an expression cassette comprising a second heterologous polynucleotide into said CHO cell by replacing the expression cassette comprising the first heterologous polynucleotide of step aa).

[024] In yet another aspect, the present disclosure refers to a method for the production of a protein of interest in a CHO cell comprising (a) providing the CHO cell of the invention; (b) culturing the CHO cell of step a) in a cell culture medium at conditions allowing production of the protein of interest; (c) harvesting the protein of interest, and (d) optionally purifying the protein of interest.

[025] The CHO cell used in the methods according to the invention may be a CHO-DG44 cell, a CHO-K1 cell, a CHO-DXB11 cell, a CHO-S cell, a CHO glutamine synthetase (GS)-deficient cell or a derivative of any of these cells.

[026] In yet another aspect, the present disclosure refers to a use of the CHO cell as disclosed herein for producing a protein of interest at high yield is described.

DESCRIPTION OF THE FIGURES

FIGURE 1: Random versus targeted integration via ZFN (Pool data) in CHO cells. (A) Shown are IgG1 antibody concentrations from randomly integrated (black bars) versus targeted integrated (white bars) CHO-DG44 cell pools after 3-7 days of fed-batch culture. (B) Shown are IgG1 antibody concentrations from randomly integrated (black bars) versus targeted integrated (white bars) CHOZN GS cell pools after 8-10 days of fed-batch culture. TI pools were enriched using FACS cell sorting, metabolic selections and a second round of FACS. Targeted integration was zinc finger nuclease (ZFN) mediated using zinc finger nuclease pair (ZFN) 13 designed to integrate downstream of the S100A3/A4/A5/A6 main gene cluster.

FIGURE 2: Productivity assessment of independent single CHOZN GS clones for homogeneity of antibody production following (A) targeted integration via ZFN or (B) random integration. Shown are 20 - 24 independent clones, which were obtained via limiting dilution following the respective transfection protocol (TI or RI, respectively). Cells were passaged over 60 days in TTP tubes. The bars represent pooled data from IgG titers in µg/ml of individual clones in fed-batch cultures after 8 days following 0 (n = 2) and 60 days (n = 2) of passaging. Error bars indicate stability of clones passaged for 0 to 60 days. Targeted integration downstream of the S100A3/A4/A5/A6 main gene cluster using ZNF 13 resulted in more homogenous clonal IgG expression levels and more stable expression over 60 days in culture of the single clones.

FIGURE 3: Effect of integration site on antibody pool titers after TI. (A) Illustration of the location of individual ZFNs and hot spot loci in the S100A gene cluster. Numbers indicate boundaries based on the *Cricetulus griseus* scaffold of CHOZN GS cells having the NCBI Reference Sequence: NW_003613854.1. The arrows indicate the integration site of ZNFs 7 to 14 and are classified into

"non disruptive and productive" (black), "non disruptive and low/non-productive" (white) and "disruptive and low/non-productive" (shaded). (B) IgG titers in mg/l are shown for CHO pools obtained using ZNFs 7 to 14 mediating integration into different loci as indicated on the X-axis.

FIGURE 4: Targeted integration via landing pad in CHO-K1 GS cells. (A) Schematic illustration of a DNA construct integrated into the CHO genome via ZFN for site specific integration of a landing pad for ZFN locus 13 (SEQ ID NO: 11) comprising homology arms (SEQ ID NOs: 13 and 14), flippase recognition target (FRT) sites FRT and FRT5 and two selection markers separated by an IRES sequence. (B) Shown are IgG1 antibody concentrations from targeted integrated CHOZN GS cell pools.

FIGURE 5: Productivity assessment of independent CHO-K1 GS single clones for antibody production following targeted integration via landing pad. Shown are IgG antibody concentrations of 10 independent single clones (black bars) and IgG antibody concentration (shaded) of the cell pool.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[027] The general embodiments "comprising" or "comprised" encompass the more specific embodiment "consisting of". Furthermore, singular and plural forms are not used in a limiting way. As used herein, the singular forms "a", "an" and "the" designate both the singular and the plural, unless expressly stated to designate the singular only.

[028] The term "homologue" or "homologous" as used in the present invention means a polypeptide molecule or a nucleic acid molecule, which is at least 80% identical in sequence with the original sequence or its complementary sequence. Preferably, the polypeptide molecule or nucleic acid molecule is at least 90% identical in sequence with the reference sequence or its complementary sequence. More preferably, the polypeptide molecule or nucleic acid molecule is at least 95% identical in sequence with the reference sequence or its complementary sequence. Most preferably, the polypeptide molecule or a nucleic acid molecule is at least 98% identical in sequence with the reference sequence or its complementary sequence. A homologous protein further displays the same or a similar protein activity as the original sequence.

[029] The term "corresponding to the sequence" or "corresponds to the sequence", as used herein includes the defined sequence of *Cricetulus griseus* CHO-K1 having the sequence or the sequence between the defined nucleotides of SEQ ID NO: 1 or SEQ ID NO: 2, but also natural variations thereof. The skilled person will understand that genomic sequences of CHO cell lines vary and may therefore not be identical with the sequences obtained from NCBI database with the NCBI Reference Sequence: NW_003613854.1, and as shown in SEQ ID NOs: 1 and 2 due to, e.g., allelic variation. However, using sequence alignment, the skilled person would know how to identify the sequence in a specific CHO cell line corresponding to the sequence as defined in SEQ ID NO: 1 or 2, i.e., the homologous region. Such corresponding sequence would have at least 80% identity with the sequence defined in SEQ ID NO: 1 or with the sequence defined in SEQ ID NO: 2, preferably at least 90% identity with the sequence defined in SEQ ID NO: 1 or with the sequence defined in SEQ

ID NO: 2 or is identical with SEQ ID NO: 1 or SEQ ID NO: 2. The corresponding sequence may also contain recombinant insertions, such as a heterologous polynucleotide, which is not to be considered for determining the corresponding sequence.

[030] The term “protein” is used interchangeably with “amino acid residue sequence” or “polypeptide” and refers to polymers of amino acids of any length. These terms also include proteins that are post-translationally modified through reactions that include, but are not limited to, glycosylation, acetylation, phosphorylation, glycation or protein processing. Modifications and changes, for example fusions to other proteins, amino acid sequence substitutions, deletions or insertions, can be made in the structure of a polypeptide while the molecule maintains its biological functional activity. For example certain amino acid sequence substitutions can be made in a polypeptide or its underlying nucleic acid coding sequence and a protein can be obtained with the same properties. The term “polypeptide” typically refers to a sequence with more than 10 amino acids and the term “peptide” means sequences with up to 10 amino acids in length. However, the terms may be used interchangeably. The protein of interest according to the present invention is preferably a therapeutic protein.

[031] The term “protein of interest” broadly refers to any protein that is of specific relevance in an industrial protein production process. Proteins of interest include, but are not limited to heterologous therapeutic proteins, marker proteins or proteins of the host cell having a function in e.g. protein secretion, post-translational protein modification, translation, transcription, cell cycle regulation or nutrient metabolism.

[032] The term “therapeutic protein” refers to proteins that can be used in medical treatment of humans and/or animals. These include, but are not limited to antibodies, growth factors, blood coagulation factors, vaccines, interferons, hormones and fusion proteins.

[033] The term "genomic DNA", or "genome" is used interchangeably and refers to the heritable genetic information of a host organism. The genomic DNA comprises the DNA of the nucleus (also referred to as chromosomal DNA) but also of other cellular organelles (e.g., mitochondria).

[034] The term "gene" as used herein refers to a DNA or RNA locus of heritable genomic sequence which affects an organism’s traits by being expressed as a functional product or by regulation of gene expression. Genes and polynucleotides may include introns and exons as in genomic sequence, or just the coding sequences as in cDNAs, such as an open reading frame (ORF), comprising a start codon (methionine codon) and a translation stop codon. Genes and polynucleotides can also include regions that regulate their expression, such as transcription initiation, translation and transcription termination. Thus, also included are regulatory elements such as a promoter.

[035] The terms "nucleic acid", "nucleotide", and "polynucleotide" as used herein are used interchangeably and refer to a single or double- stranded polymer of deoxyribonucleotide bases or ribonucleotide bases read from the 5' to the 3' end and include double stranded DNA (dsDNA), single stranded DNA (ssDNA), single stranded RNA (ssRNA), double stranded RNA (dsRNA),

genomic DNA, cDNA, cRNA, recombinant DNA or recombinant RNA and derivatives thereof, such as those containing modified backbones. Preferably, a polynucleotide, particularly to be stably integrated into the CHO genome is a DNA or cDNA. Polynucleotides according to the invention can be prepared in different ways (e.g. by chemical synthesis, by gene cloning etc.) and can take various forms (e.g. linear or branched, single or double stranded, or a hybrid thereof, primers, probes etc.). The term "nucleotide sequence" or "nucleic acid sequence" refers to both the sense and antisense strands of a nucleic acid as either individual single strands or in the duplex.

[036] The term "heterologous polynucleotide" as used herein refers to a polynucleotide derived from a different organism or a different species from the recipient, i.e., a CHO cell. In the context of the present invention the skilled person would understand that it refers to a DNA or cDNA. A heterologous polynucleotide may also be referred to as transgene. Thus, it may be a gene or open reading frame (ORF) coding for a heterologous protein. In the context of the CHO cell “heterologous polynucleotide” refers to a polynucleotide derived from a different cell line, preferably a cell line not derived from *Cricetulus griseus*. The term "heterologous" when used with reference to portions of a nucleic acid may also indicate that the nucleic acid comprises two or more sequences that are not found in the same relationship to each other in nature. Heterologous may therefore also refer to a CHO derived polynucleotide sequence, such as a gene or transgene, or a portion thereof, being inserted into the CHO genome in a location in which it is not typically found, or a gene introduced into a cell of an organism in which it is not typically found.

[037] “Heterologous polynucleotide”, “heterologous gene” or “heterologous sequences” can be introduced into a target cell directly or preferably by using an “expression vector”, preferably a mammalian expression vector. Methods used to construct vectors are well known to the person skilled in the art and described in various publications. In particular techniques for constructing suitable vectors, including a description of the functional components such as promoters, enhancers, termination and polyadenylation signals, selection markers, origins of replication, and splicing signals, are reviewed in considerable details in (Sambrook J, *et al.*, 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press) and references cited therein. Vectors may include but are not limited to plasmid vectors, phagemids, cosmids, artificial/mini-chromosomes (e.g. ACE), or viral vectors such as baculovirus, retrovirus, adenovirus, adeno-associated virus, herpes simplex virus, retroviruses and bacteriophages. The eukaryotic expression vectors will typically contain also prokaryotic sequences that facilitate the propagation of the vector in bacteria such as an origin of replication and antibiotic resistance genes for selection in bacteria. A variety of eukaryotic expression vectors, containing a cloning site into which a polynucleotide can be operably linked, are well known in the art and some are commercially available from companies such as Stratagene, La Jolla, CA; Invitrogen, Carlsbad, CA; Promega, Madison, WI or BD Biosciences Clonetech, Palo Alto, CA. Usually expression vectors also comprise an expression cassette encoding a selectable marker, allowing selection of host cells carrying said expression marker.

[038] The term “producing” or “highly producing”, “production”, “production and/or secretion”, “producing”, “production cell” or “producing at high yield” as used herein relates to the production of the RNA and/or protein encoded by a heterologous polynucleotide. An “increased production and/or secretion” or “production at high yield” relates to the expression of the heterologous RNA and/or protein and means an increase in specific productivity, increased titer, increased overall productivity of the cell culture or a combination thereof. Preferably, the titer or the overall productivity and the titer are increased. Increased titer as used herein relates to an increased concentration in the same volume, i.e., an increase in total yield. The produced heterologous RNA, heterologous protein or therapeutic protein may be, for example, a small regulatory RNA or an antibody, preferably a micro RNA, a small hairpin RNA, a monoclonal antibody, a bispecific antibody or a fragment thereof, or a fusion protein.

[039] The term “enhancement”, “enhanced”, “enhanced”, “increase” or “increased”, as used herein, generally means an increase by at least about 10% as compared to a control cell, for example an increase by at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 100%, or at least about 200%, or at least about 300%, or any integer decrease between 10-300% as compared to a control cell. As used herein, a “control cell” or “control mammalian cell” is the same CHO cell in which the same heterologous polynucleotide has been introduced randomly. This may be determined in cell clones or preferably in a cell pool without clonal selection.

[040] As used herein, the term “expression cassette” refers to the part of a vector comprising one or more genes encoding for a RNA (heterologous RNA) or a protein (heterologous protein) and the sequences controlling their expression. Thus it comprises a promoter sequence, an open reading frame and a 3’ untranslated region, typically containing a polyadenylation site. Preferably, the vector is an expression vector comprising one or more gene encoding for the recombinant secreted therapeutic protein. It may be part of a vector, typically an expression vector, including a plasmid or a viral vector. It may also be integrated into a chromosome by random or targeted integration, such as by homologous recombination. An expression cassette is prepared using cloning techniques and does therefore not refer to a natural occurring gene structure.

[041] A “promoter” or “promoter sequence” is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3′ direction) coding sequence. The promoter sequence is bounded at its 3′ terminus by the transcription initiation site and extends upstream (5′ direction) to include up to 1.5 kb. Typically, a promoter is about 100 to 1000 base pairs long. A promoter sequence comprises a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters often, but not always, contain “TATA” boxes and “CAT” boxes. Promoter sequences often contain additional consensus sequences recognized by proteins involved in regulating expression of the respective gene. Regulation of gene expression by a promoter can occur by enhancing or inhibiting binding of a regulatory protein. Enhancing or inhibiting the binding or a regulatory protein

can occur by many different means, including but not limited to, base modifications (i.e., methylation) and protein modification (i.e., phosphorylation).

[042] The terms “upstream” and “downstream” refer to a relative position in DNA or RNA. Each strand of DNA or RNA possesses a 5’ end and a 3’ end, relating to the terminal carbon position of the deoxyribose or ribose units. By convention, “upstream” means towards the 5’ end of a polynucleotide, whereas “downstream” means towards the 3’ end of a polynucleotide. In the case of double stranded DNA, e.g. genomic DNA, the term “upstream” means towards the 5’ end of the coding strand, whereas “downstream” means towards the 3’ end of the coding strand.

[043] The term “coding strand”, “sense strand” or “non-template strand” refers to the strand of the double stranded DNA whose base sequence corresponds to the base sequence of the RNA which is transcribed from a gene.

[044] The term “small regulatory RNA” refers to small non-coding RNA polynucleotides that influence the expression of target genes, usually by binding to their respective mRNAs. These small regulatory RNAs include, but are not limited to small interfering RNAs (siRNAs), micro RNAs (miRNAs) and short hairpin RNAs (shRNAs).

[045] The term “ribonucleic acid”, “RNA” or “RNA oligonucleotide” as used herein describes a molecule consisting of a sequence of nucleotides, which are built of a nucleobase, a ribose sugar, and a phosphate group. RNAs are usually single stranded molecules and can exert various functions. The term ribonucleic acid specifically comprises messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), short interfering RNA (siRNA), small hairpin RNA (shRNA) and micro RNA (miRNA), each of which plays a specific role in biological cells. It includes small non-coding RNAs, such as microRNAs (miRNA), short interfering RNAs (siRNA), small hairpin RNA (shRNA), and Piwi-interacting RNAs (piRNA). The term "non-coding" means that the RNA molecule is not translated into an amino acid sequence.

[046] The term “RNA interference” (RNAi) refers to sequence-specific or gene-specific suppression of gene expression (protein synthesis), without generalized suppression of protein synthesis. RNAi may involve degradation of messenger RNA (mRNA) by an RNA-induced silencing complex (RISC), preventing translation of the transcribed mRNA. The suppression of gene expression caused by RNAi may be transient or it may be more stable, even permanent. RNAi may be mediated by miRNA, siRNA or shRNA. Preferably the RNAi according to the invention is gene-specific (only one gene is targeted). Gene-specific RNAi may be mediated by siRNA or shRNA.

[047] The terms “microRNA” or “miRNA” are used interchangeably herein. microRNAs are small, about 22 nucleotide-long (typically between 19 and 25 nucleotides in length) non-coding single stranded RNAs. miRNAs typically target more than one gene. microRNAs are encoded in the genome of eukaryotic cells and are typically transcribed by RNA Polymerase III as long primary transcripts that are then processed in several steps first into ~70nt-long hairpin-loop structures and subsequently into the ~22nt RNA duplex. The active mature strand is then loaded into the RNA-induced silencing complex (RISC) in order to block translation of target proteins or degradation of

their respective mRNAs. Targeting with miRNAs allows for mismatches and mRNA translational repression is mediated by incomplete complementarity (i.e., imperfect base paring between the antisense strand of the RNA duplex of the small interfering RNA and the target mRNA), while siRNA and shRNA are specific for their targets due to complete sequence complementarity (i.e., perfect base pairing between the antisense strand of the RNA duplex of the small interfering RNA and the target mRNA). Typically, miRNAs bind in the 3'untranslated region (3'UTR) and are not gene-specific, but target multiple mRNAs. The term "microRNA" as used herein relates to endogenous genomic mammalian miRNAs, such as human miRNAs. The prefix "hsa" indicates, e.g., the human origin of a microRNA. They may be introduced into a mammalian host cell using an expression vector comprising genomic microRNA sequence(s) for transient or stable expression of miRNA in the mammalian host cell. Means for cloning genomic microRNA into an expression vector are known in the art. They include, cloning genomic miRNA sequences with approximately 300 bp flanking regions into a mammalian expression vector, such as pBIP-1, operably linked to a promoter. Alternatively one or more microRNAs may be cloned as polynucleotides encoding engineered pre-miRNA sequences (i.e., short hairpins) into a mammalian expression vector. For example, a mature miRNA sequence may be cloned into a given sequence encoding an optimized hairpin loop sequence and 3' and 5' flanking regions, such as derived from the murine miRNA mir-155 (Lagos-Quintana *et al.*, 2002. Curr. Biol. 30;12(9):735-9). A DNA oligonucleotide is designed, which encodes the miRNA sequence, the mentioned loop and the antisense sequence of the respective mature miRNA with a two nucleotide depletion to generate an internal loop in the hairpin stem. Furthermore, overhangs are added for cloning at both ends to fuse the DNA oligonucleotide to the 3' and 5' flanking regions. miRNAs as used herein further comprise non-canonical miRNAs. These RNAs can be derived from 'housekeeping' non-coding RNAs (ncRNA) including ribosomal RNA (rRNA) or transfer RNA (tRNA) and function in a miRNA-like manner. These RNAs can also originate from mammalian mitochondrial ncRNAs and are termed mitochondrial genome-encoded small RNAs (mitosRNAs).

[048] As used herein, the terms "small interfering" or "short interfering RNA" or "siRNA" refer to an RNA duplex of nucleotides that is targeted to a desired gene and is capable of inhibiting the expression of a gene with which it shares homology. It is formed from long double stranded RNA (dsRNA) or shRNA. The RNA duplex typically comprises two complementary single-stranded RNAs of 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 or 29 nucleotides that form 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 or 27 base pairs and possess 3′ overhangs of two nucleotides, preferably the RNA duplex comprises two complementary single stranded RNAs of 19-27 nucleotides that form 17-25 base pairs and possess 3' overhangs of two nucleotides. siRNA is "targeted" to a gene, wherein the nucleotide sequence of the duplex portion of the siRNA is complementary to a nucleotide sequence of the mRNA of the targeted gene. The siRNA or a precursor thereof is always exogenously introduced into the cell, e.g., directly or by transfection of a vector having a sequence encoding said siRNA, and the endogenous miRNA pathway is harnessed for correct processing of siRNA and cleavage or degradation of the target mRNA. The duplex RNA can be expressed in a cell from a single construct.

[049] As used herein, the term “shRNA” (small hairpin RNA) refers to an RNA duplex wherein a portion of the siRNA is part of a hairpin structure (shRNA). The shRNA can be processed intracellularly into a functional siRNA. In addition to the duplex portion, the hairpin structure may contain a loop portion positioned between the two sequences that form the duplex. The loop can vary in length. In some embodiments the loop is 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 nucleotides in length. The hairpin structure can also contain 3′ or 5′ overhang portions. In some aspects, the overhang is a 3′ or a 5′ overhang of 0, 1, 2, 3, 4 or 5 nucleotides in length. In one aspect of this invention, a nucleotide sequence comprised in the vector serves as a template for the expression of a small hairpin RNA, comprising a sense region, a loop region and an antisense region. Following expression the sense and antisense regions form a duplex. shRNA is always exogenously introduced, e.g., by transfection of a vector having a sequence encoding said shRNA, and the endogenous miRNA pathway is harnessed for correct processing of the siRNA and cleavage or degradation of the target mRNA. Use of a vector having a sequence encoding a shRNA has the advantage over use of chemically synthesized siRNA in that the suppression of the target gene is typically long-term and stable.

[050] Typically siRNA and shRNA mediate mRNA repression by complete sequence complementarity (i.e., perfect base paring between the antisense strand of the RNA duplex of the small interfering RNA and the target mRNA) and are therefore specific for their target. The antisense strand of the RNA duplex may also be referred to as active strand of the RNA duplex. Complete sequence complementarity of perfect base paring as used herein means that the antisense strand of the RNA duplex of the small interfering RNA has at least 89% sequence identity with the target mRNA for at least 15 continuous nucleotides, at least 16 continuous nucleotides, at least 17 continuous nucleotides, at least 18 continuous nucleotides and preferably at least 19 continuous nucleotides, or preferably at least 93% sequence identity with the target mRNA for at least 15 continuous nucleotides, at least 16 continuous nucleotides, at least 17 continuous nucleotides, at least 18 continuous nucleotides and preferably at least 19 continuous nucleotides. More preferably the antisense strand of the RNA duplex of the small interfering RNA has 100% sequence identity with the target mRNA for at least 15 continuous nucleotides, at least 16 continuous nucleotides, at least 17 continuous nucleotides, at least 18 continuous nucleotides and preferably at least 19 continuous nucleotides.

[051] A “vector” is a nucleic acid that can be used to introduce a heterologous polynucleotide into a cell. One type of vector is a “plasmid”, which refers to a linear or circular double stranded DNA molecule into which additional nucleic acid segments can be ligated. Another type of vector is a viral vector (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), wherein additional DNA or RNA segments can be introduced into the viral genome. Preferably the vector is a non-episomal mammalian vector integrating into the genome of a host cell upon introduction into the host cell and culturing under selective pressure, and thereby are replicated along with the host genome. A vector can be used to direct the expression of a chosen polynucleotide in a cell.

[052] The term “encodes” and “codes for” refers broadly to any process whereby the information in a polymeric macromolecule is used to direct the production of a second molecule that is different from the first. The second molecule may have a chemical structure that is different from the chemical nature of the first molecule. For example, in some aspects, the term “encode” describes the process of semi-conservative DNA replication, where one strand of a double-stranded DNA molecule is used as a template to encode a newly synthesized complementary sister strand by a DNA-dependent DNA polymerase. In other aspects, a DNA molecule can encode an RNA molecule (e.g., by the process of transcription that uses a DNA-dependent RNA polymerase enzyme). Also, an RNA molecule can encode a polypeptide, as in the process of translation. When used to describe the process of translation, the term “encode” also extends to the triplet codon that encodes an amino acid. In some aspects, an RNA molecule can encode a DNA molecule, e.g., by the process of reverse transcription incorporating an RNA-dependent DNA polymerase. In another aspect, a DNA molecule can encode a polypeptide, where it is understood that “encode” as used in that case incorporates both the processes of transcription and translation.

[053] As used herein the term “gene cluster” refers to a segment of genomic DNA that encompasses a set or family of closely related genes which code for a group of related, or similar proteins and which are usually grouped together on the same chromosome. A gene cluster encompasses a segment of genomic DNA, wherein all the coding sequences for the group of proteins are located, including regions preceding (leader) and following (trailer) the coding sequences as well as intervening sequences (introns) between individual coding sequence fragments (exons) and further genetic elements in the broadest sense, including, but not limited to, transcriptional regulator elements, promoter elements, enhancer elements and repressor elements. Generally, the gene cluster encompasses the entire genomic segment limited by the first (5’) protein coding gene of the gene cluster and the last (3’) protein coding gene of the gene cluster.

[054] The “S100A gene cluster” refers to a segment of Chinese hamster genomic DNA that codes for the group of calcium binding proteins S100A1, S100A3, S100A4, S100A5, S100A6, S100A13, S100A14 and S100A16. The segment comprises the most upstream gene coding for the S100A1 protein and the most downstream gene coding for the S100A6 protein. The term “S100A3/A4/A5/A6 main gene cluster” refers to a segment of genomic DNA that is encompassed by the S100A gene cluster and reaches from the gene coding for the S100A3 protein to the gene coding for the S100A6 protein (SEQ ID NO: 4). “S100A1“ refers to the protein S100A1 from *Cricetulus griseus* and the gene coding for it (the S100A1 gene; NCBI Gene ID: 100769478). “S100A3” refers to the protein S100A3 from *Cricetulus griseus* and the gene coding for it (the S100A3 gene, NCBI Gene ID: 100770814). “S100A4” refers to the protein S100A4 from *Cricetulus griseus* and the gene coding for it (the S100A4 gene, NCBI Gene ID: 100770532). “S100A5” refers to the protein S100A5 from *Cricetulus griseus* and the gene coding for it (the S100A5 gene, NCBI Gene ID: 100771097). “S100A6” refers to the protein S100A6 from *Cricetulus griseus* and the gene coding for it (the S100A6 gene; NCBI Gene ID: 100771384). “S100A13” refers to the protein S100A13 from *Cricetulus griseus* and the gene coding for it (the S100A13 gene; NCBI Gene ID: 100769763). “S100A14” refers to the

protein S100A14 from *Cricetulus griseus* and the gene coding for it (the S100A14 gene; NCBI Gene ID: 100770053). “S100A16” refers to the protein S100A16 from *Cricetulus griseus* and the gene coding for it (the S100A16 gene; NCBI Gene ID: 100753026).

[055] The term “allele” refers to any one of the different forms of a gene, genetic target region or generally DNA sequence at a single locus, i.e., chromosomal location. This includes coding sequences, non-coding sequences and regulatory sequences. Different alleles within a genome are not necessarily identical in nucleotide sequence.

[056] The term “antibody” refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant regions genes as well as the myriad immunoglobulin variable region genes. The terms “antibody” and “immunoglobulin” are used interchangeably and are used to denote, without being limited thereto, glycoproteins having the structural characteristics noted above for immunoglobulins.

[057] The term “antibody” is used herein in its broadest sense and encompasses monoclonal antibodies (including full length monoclonal antibodies), polyclonal antibodies, chimeric antibodies, humanized antibodies, human antibodies, multispecific antibodies (e.g. bispecific antibodies), single domain antibodies, and antibody fragments (such as Fv, Fab, Fab´, F(ab)2 or other antigen-binding subsequences of antibodies). The term “antibody” also encompasses antibody conjugates and fusion antibodies. Full length “antibodies” or “immunoglobulins” are generally heterotetrameric glycoproteins of about 150 kDa, composed of two identical light and two identical heavy chains. Each light chain is linked to a heavy chain by one covalent disulphide bond, while the number of disulphide linkages varies between the heavy chains of different immunoglobulin isotypes. Each heavy and light chain also has regularly spaced intrachain disulphide bridges. Each heavy chain has an amino terminal variable domain (VH) followed by three carboxy terminal constant domains (CH). Each light chain has a variable N-terminal domain (VL) and a single C-terminal constant domain (CL). The term “antibody” further refers to a type of antibody comprising a plurality of individual antibodies having the same specificity (variable domain) and having the same constant domains.

[058] A “fusion protein” is defined as a protein which contains the complete sequences or any parts of the sequences of two or more originally separate natural or modified heterologous proteins or a composition of complete sequences or any parts of the sequences of two or more originally separate natural or modified heterologous proteins. Fusion proteins can be constructed by genetic engineering approaches by fusing the two or more genes, or parts thereof, that originally encode the two or more originally separate natural or heterologous proteins, or parts thereof. This results in a fusion protein with functional properties derived from each of the original proteins. Fusion proteins include, but are not limited to Fc fusion proteins.

[059] The term “cytokine” refers to small proteins, which are released by cells and act as intercellular mediators, for example influencing the behavior of the cells surrounding the secreting cell. Cytokines may be secreted by immune or other cells, such as T-cells, B-cells, NK cells and

macrophages. Cytokines may be involved in intercellular signaling events, such as autocrine signaling, paracrine signaling and endocrine signaling. They may mediate a range of biological processes including, but not limited to immunity, inflammation, and hematopoiesis. Cytokines may be chemokines, interferons, interleukins, lymphokines or tumor necrosis factors.

[060] As used herein, "growth factor" refers to proteins or polypeptides that are capable of stimulating cell growth. They include, but are not limited to, insulin, epidermal growth factor (EGF), ephrins (Eph), Erythropoietin, glia-cell stimulating factor (GSF); colony-stimulating factors (CSF) including macrophage colony-stimulating factor (M-CSF), granulocyte macrophage colony-stimulating factor (GM-CSF), and granulocyte colony-stimulating factor (G-CSF); stem cell growth factor (SCGF) (also called Steel Factor); stromal cell-derived factor (SDF), effective fragments thereof, and combinations thereof; and vascular endothelial growth factor (VEGF). Other growth factors can include hepatocyte growth factor (HGF), Angiopoietin-1, Angiopoietin-2, b-FGF, and FLT-3 ligand, and effective fragment thereof.

[061] The term "expression" as used herein refers to transcription and/or translation of a heterologous nucleic acid sequence within a host cell. The level of expression of a gene product of interest in a host cell may be determined on the basis of either the amount of corresponding RNA that is present in the cell, or the amount of the polypeptide encoded by the selected sequence. For example, RNA transcribed from a selected sequence can be quantified by Northern blot hybridization, ribonuclease RNA protection, in situ hybridization to cellular RNA or by PCR, such as qPCR. Proteins encoded by a selected sequence can be quantitated by various methods, e.g. by ELISA, by Western blotting, by radioimmunoassay, by immunoprecipitation, by assaying for the biological activity of the protein, by immunostaining of the protein followed by FACS analysis or by homogeneous time-resolved fluorescence (HTRF) assays. The level of expression of a non-coding RNA, such as a miRNA or shRNA may be quantified by PCR, such as qPCR.

[062] The term "gene product" refers to both the RNA polynucleotide and polypeptide that is encoded by a gene or DNA polynucleotide.

[063] A "marker gene" as used herein means a polynucleotide, the expression of which in a cell confers a selectable or distinguishable phenotype (e.g., antibiotic resistance, expression of a fluorescent protein or reporter gene, modified metabolism) to the cell.

[064] As used herein, a "reporter gene" is a polynucleotide encoding a protein whose expression by a host cell can be detected and quantified. Thus, a measurement of the level of expression of the reporter is typically indicative of the level of activation of the promoter element that directs expression of the gene encoding the reporter (reporter gene) within the host cell genome. For example, a reporter gene can encode a protein, for example, an enzyme whose activity can be quantified, for example, alkaline phosphatase (AP), chloramphenicol acetyltransferase (CAT), Renilla luciferase or firefly luciferase protein(s). Reporters also include fluorescent proteins, for example, green fluorescent protein (GFP) or any of the recombinant variants of GFP, including enhanced GFP (EGFP), blue fluorescent proteins (BFP and other derivatives), cyan fluorescent

protein (CFP and other derivatives), yellow fluorescent protein (YFP and other derivatives) and red fluorescent protein (RFP and other derivatives).

[065] A “selectable marker gene” or “selection marker gene” is a gene which encodes a selectable marker and allows the specific selection of cells which contain this gene, typically by the addition of a corresponding “selecting agent” to the cultivation medium. As an illustration, an antibiotic resistance gene may be used as a positive selectable marker. Only cells which have been transformed with this gene are able to grow in the presence of the corresponding antibiotic and are thus selected. Untransformed cells, on the other hand, are unable to grow or survive under these selection conditions. There are positive, negative and bifunctional selectable markers. Positive selectable markers permit the selection and hence enrichment of transformed cells by conferring resistance to the selecting agent or by compensating for a metabolic or catabolic defect in the host cell. By contrast, cells which have received the gene for the selectable marker can be selectively eliminated by negative selectable markers. An example of this is the thymidine kinase gene of the Herpes Simplex virus, the expression of which in cells with the simultaneous addition of acyclovir or ganciclovir leads to the elimination thereof. The selectable marker genes useful in this invention also include the amplifiable selectable markers. The literature describes a large number of selectable marker genes including bifunctional (positive/negative) markers (see for example WO 92/08796 and WO 94/28143). Examples of selectable markers which are useful in the present invention include, but are not limited to the genes of aminoglycoside phosphotransferase (APH), hygromycine phosphotransferase (HYG), dihydrofolate reductase (DHFR), thymidine kinase (TK), glutamine synthetase, asparagine synthetase and genes which confer resistance to neomycin (G418/Geneticin), puromycin, histidinol D, bleomycin, phleomycin, blasticidin and zeocin. Also included are genetically modified mutants and variants, fragments, functional equivalents, derivatives, homologues and fusions with other proteins or peptides, provided that the selectable marker retains its selective qualities. Such derivatives display considerable homology in the amino acid sequence in the regions or domains, which are deemed to be selective.

[066] Selection may also be made by fluorescence activated cell sorting (FACS) using for example a cell surface marker, bacterial β-galactosidase or fluorescent proteins (e.g. green fluorescent proteins (GFP) and their variants from Aequorea victoria and Renilla reniformis or other species; red fluorescent proteins, fluorescent proteins and their variants from non-bioluminescent species (e.g. Discosoma sp., Anemonia sp., Clavularia sp., Zoanthus sp.) to select for recombinant cells.

[067] The term “selection agent” or “selective agent” refers to a substance that interferes with the growth or survival of a cell, unless a certain selectable marker gene product is present in the cell which alleviates the effect of the selection agent. For example, to select for the presence of an antibiotic resistance gene like APH (aminoglycoside phosphotransferase) in a transfected cell the antibiotic Geneticin (G418) is used.

[068] The “amplifiable selectable marker gene” usually codes for an enzyme, which is needed for the growth of eukaryotic cells under certain cultivation conditions. For example, the amplifiable selectable marker gene may code for dihydrofolate reductase (DHFR) or glutamine synthetase (GS).

In this case the marker gene is amplified, if a host cell transfected therewith is cultivated in the presence of the selecting agent methotrexate (MTX) or methionine sulphoximine (MSX), respectively. Sequences linked to the amplifiable selectable marker gene (i.e., sequences physically proximal thereto) are co-amplified together with the amplifiable selectable marker gene. Said co-amplified sequences may be introduced on the same expression vector or on separate vectors.

[069] The following Table 1 gives non-limiting examples of amplifiable selectable marker genes and the associated selecting agents, which may be used according to the invention. Suitable amplifiable selectable marker genes are also described in an overview by Kaufman (Kaufman RJ, 1990. Methods Enzymol. 185:537-566).

Table 1: Amplifiable selectable marker genes

Amplifiable selectable marker gene	Accession number	Selecting agent
dihydrofolate reductase (DHFR)	M19869 (hamster) E00236 (mouse)	methotrexate (MTX)
metallothionein	D10551 (hamster) M13003 (human) M11794 (rat)	cadmium
CAD (carbamoylphosphate synthetase : aspartate transcarbamylase: dihydroorotase)	M23652 (hamster) D78586 (human)	N-phosphoacetyl-L-aspartate
adenosine-deaminase	K02567 (human) M10319 (mouse)	Xyl-A- or adenosine, 2´deoxycoformycin
AMP (adenylate)-deaminase	D12775 (human) J02811 (rat)	adenine, azaserin, coformycin
UMP-synthase	J03626 (human)	6-azauridine, pyrazofuran
IMP 5´-dehydrogenase	J04209 (hamster) J04208 (human) M33934 (mouse)	mycophenolic acid
xanthine-guanine-phosphoribosyltransferase	X00221 (E. coli)	mycophenolic acid with limiting xanthine
mutant HGPRTase or mutant thymidine-kinase	J00060 (hamster) M13542, K02581 (human) J00423, M68489(mouse) M63983 (rat) M36160 (Herpes virus)	hypoxanthine, aminopterine and thymidine (HAT)
thymidylate-synthetase	D00596 (human) M13019 (mouse) L12138 (rat)	5-fluorodeoxyuridine
P-glycoprotein 170 (MDR1)	AF016535 (human) J03398 (mouse)	several drugs, e.g. adriamycin, vincristin, colchicine
ribonucleotide reductase	M124223, K02927 (mouse)	aphidicoline
glutamine-synthetase (GS)	AF150961 (hamster) U09114, M60803 (mouse) M29579 (rat)	methionine sulphoximine (MSX)
asparagine-synthetase	M27838 (hamster) M27396 (human) U38940 (mouse) U07202 (rat)	β-aspartylhydroxamate, albizziin, 5´azacytidine
argininosuccinate-synthetase	X01630 (human) M31690 (mouse) M26198 (bovine)	canavanin
ornithine-decarboxylase	M34158 (human)	α-difluoromethylornithine

	J03733 (mouse) M16982 (rat)	
HMG-CoA-reductase	L00183,M12705 (hamster) M11058 (human)	compactin
N-acetylglucosaminyl-transferase	M55621 (human)	tunicamycin
threonyl-tRNA-synthetase	M63180 (human)	borrelidin
Na^+K^+-ATPase	J05096 (human) M14511 (rat)	ouabain

[070] According to the invention a preferred amplifiable selectable marker gene is a gene which codes for a polypeptide with the function of GS or DHFR.

[071] The term “site specific recombinase” refers to proteins that recognize specific nucleotide sequences (recognition sites), cleave the DNA backbone at these sites, perform a rearrangement and re-ligate the cleaved nucleotide sequences. Said recombinases for example allow the excision of the DNA between a pair of recognition sites and the subsequent integration of a polynucleotide of interest instead of the excised DNA fragment, thereby providing a precise site-specific exchange of genetic information. Several site-specific recombinases are known in the art. For instance, Cre recombinase recognizes either loxP recombination sites or lox511 recombination sites which are hetero-specific, which means that loxP and lox511 do not recombine together. The Cre/lox system is, e.g., described in Odell *et al.*, Plant Physiol. 1994, 106(2), 447-58. Flp recombinase recognizes frt recombination sites as , e.g., described in Lyznik *et al.*, Nucleic Acids Res. 1996, 24(19), 3784-9. The phiC31 integrase recognizes attachment (att) sites, such as attB (donor) and attP (acceptor) as, e.g., described in Groth *et al.*, Proc. Natl. Acad. Sci. U. S. A. 2000, 97(11), 5995-6000. The Dre recombinase recognizes rox sites as, e.g., described in U.S. Pat. No. 7,422,889. The Int recombinase from bacteriophage lambda (lambda integrase) and its recombination sites are described in Landy, Annu. Rev. Biochem. 1989, 58, 913-49.

[072] According to the invention, a “sequence specific DNA editing enzyme” or a “site specific nuclease” is a protein that enables the cleavage of DNA at defined nucleotide sequences (recognition sites). Said cleavage may occur on one or both of two complementary DNA strands and thus allow, for example targeted mutagenesis, targeted deletion of specific genomic DNA sequences or result in the site-directed recombination of the cleaved target DNA with a heterologous polynucleotide. The sequence specificity of said editing enzymes may result from one or more sequence specific DNA binding protein domains within the editing enzyme, or from the enzyme binding a guide polynucleotide (e.g. guide RNA) that directs it to a DNA sequence with at least partial complementarity to said guide polynucleotide. The recognition site of said editing enzymes may therefore be altered by engineering the DNA binding protein domains, or using alternative guide polynucleotides. Multiple sequence specific DNA editing enzymes are known in the art, non-limiting examples of which are zinc finger nucleases (ZFNs), meganucleases, transcription activator-like effector nucleases (TALENs) and CRISPR associated nucleases.

[073] The term “stable integration” or “stably integrated” as used in the patent refers to a heterologous polynucleotide being introduced into a host cell genome, as opposed to transiently

introduced polynucleotides that remain separate from the genomic DNA of the host cell. Stable integration may occur by homologous recombination or other types of recombination. Stable integration may comprise a step of transient introduction of a heterologous polynucleotide into a host cell.

Stable integration of at least one heterologous polynucleotide into the S100A gene cluster

[074] The present invention relates to a CHO cell comprising at least one heterologous polynucleotide, stably integrated into the S100A gene cluster of the CHO cell genome, wherein (a) the at least one heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1 (referred to as upstream genomic target region); and/or (b) the at least one heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2 (referred to as downstream genomic target region).

[075] The S100A3/A4/A5/A6 main gene cluster refers to the genomic region encompassing the Chinese hamster genes coding for the S100 calcium binding protein A3 (S100A3), the S100 calcium binding protein A4 (S100A4), the S100 calcium binding protein A5 (S100A5) and the S100 calcium binding protein A6 (S100A6) in the above order, i.e., the region from the start of S100A3 to the end of S100A6 (corresponding to 1,782,882 to 1,810,338 of *Cricetulus griseus* unplaced genomic scaffold, CriGri_1.0 scaffold682, whole genome shotgun sequence of the CHO-K1 cell line; NCBI Reference Sequence: NW_003613854.1, corresponding to the sequence of SEQ ID NO: 4, or a homologous thereof). The genomic target region upstream of the S100A3/A4/A5/A6 main gene cluster refers to a genomic region corresponding to the sequence of SEQ ID NO: 1. The genomic target region downstream of the S100A3/A4/A5/A6 main gene cluster refers to a genomic region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2.

[076] In one embodiment the at least one heterologous polynucleotide is stably integrated into the upstream genomic target region corresponding to nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1; preferably into the upstream genomic target region corresponding to nucleotides 11,820 to 18,720 of SEQ ID NO: 1, nucleotides 13,560 to 18,720 of SEQ ID NO: 1, nucleotides 15,360 to 18,720 of SEQ ID NO: 1 or nucleotides 17,100 to 18,720 of SEQ ID NO: 1, and more preferably into the upstream genomic target region corresponding to nucleotides 11,820 to 18,380 of SEQ ID NO: 1, nucleotides 13,560 to 18,380 of SEQ ID NO: 1, nucleotides 15,360 to 18,380 of SEQ ID NO: 1 or nucleotides 17,100 to 18,380 of SEQ ID NO: 1.

[077] In another embodiment the at least one heterologous polynucleotide is stably integrated into the downstream genomic target region corresponding to nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2, preferably into

the downstream genomic target region corresponding to nucleotides 660 to 10,260 of SEQ ID NO: 2, nucleotides 1,320 to 10,260 of SEQ ID NO: 2 or nucleotides 1,480 to 10,260 of SEQ ID NO: 2; and more preferably into the downstream genomic target region corresponding to nucleotides 3,180 to 10,260 of SEQ ID NO: 2, nucleotides 4,920 to 9,000 of SEQ ID NO: 2 or nucleotides 6,720 to 8,460 of SEQ ID NO: 2.

[078] In another embodiment the at least one heterologous polynucleotide is stably integrated into the upstream genomic target region and into the downstream genomic target region as disclosed above. Wherein the at least one heterologous polynucleotide integrated into the upstream genomic target region and the at least one heterologous polynucleotide stably integrated into the downstream, genomic target region may be the same or different.

[079] The skilled person will understand that a single copy, a plurality of copies of one heterologous polynucleotide, or two or more different heterologous polynucleotides may be stably integrated into the upstream genomic target region, into the downstream genomic target region, or into the upstream genomic target region and the downstream genomic target region.

[080] The at least one heterologous polynucleotide may be stably integrated into one or both alleles of the genomic target region(s).

[081] In another aspect the present invention relates to a method for the production of a CHO cell comprising the steps of (a) providing a CHO cell; (b) introducing a heterologous polynucleotide into said CHO cell, wherein the heterologous polynucleotide is stably integrated into the S100A gene cluster of the CHO cell genome, wherein (i) said heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or (ii) said heterologous polynucleotide is integrating downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2.

[082] In one embodiment the at least one heterologous polynucleotide is stably integrated into the upstream genomic target region corresponding to nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1; preferably into the upstream genomic target region corresponding to nucleotides 11,820 to 18,720 of SEQ ID NO: 1, nucleotides 13,560 to 18,720 of SEQ ID NO: 1, nucleotides 15,360 to 18,720 of SEQ ID NO: 1 or nucleotides 17,100 to 18,720 of SEQ ID NO: 1, and more preferably into the upstream genomic target region corresponding to nucleotides 11,820 to 18,380 of SEQ ID NO: 1, nucleotides 13,560 to 18,380 of SEQ ID NO: 1, nucleotides 15,360 to 18,380 of SEQ ID NO: 1 or nucleotides 17,100 to 18,380 of SEQ ID NO: 1.

[083] In another embodiment the at least one heterologous polynucleotide is stably integrated into the downstream genomic target region corresponding to nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2, preferably into

the downstream genomic target region corresponding to nucleotides 660 to 10,260 of SEQ ID NO: 2, nucleotides 1,320 to 10,260 of SEQ ID NO: 2 or nucleotides 1,480 to 10,260 of SEQ ID NO: 2; and more preferably into the downstream genomic target region corresponding to nucleotides 3,180 to 10,260 of SEQ ID NO: 2, nucleotides 4,920 to 9,000 of SEQ ID NO: 2 or nucleotides 6,720 to 8,460 of SEQ ID NO: 2.

[084] In another embodiment the heterologous polynucleotide stably integrated into the genome of the CHO cell of the invention or the CHO cell produced by the method of the invention is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region having the sequence of SEQ ID NO: 1, or at least 80% homology thereto; and/or (ii) said heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region having the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2 or at least 80% homology thereto.

[085] In one embodiment the at least one heterologous polynucleotide is stably integrated into the upstream genomic target region having the sequence of nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1, or at least 80% homology thereto; preferably into the upstream genomic target region having the sequence of nucleotides 11,820 to 18,720 of SEQ ID NO: 1, nucleotides 13,560 to 18,720 of SEQ ID NO: 1, nucleotides 15,360 to 18,720 of SEQ ID NO: 1 or nucleotides 17,100 to 18,720 of SEQ ID NO: 1, or at least 80% homology thereto, and more preferably into the upstream genomic target region having the sequence of nucleotides 11,820 to 18,380 of SEQ ID NO: 1, nucleotides 13,560 to 18,380 of SEQ ID NO: 1, nucleotides 15,360 to 18,380 of SEQ ID NO: 1 or nucleotides 17,100 to 18,380 of SEQ ID NO: 1, or at least 80% homology thereto.

[086] In another embodiment the at least one heterologous polynucleotide is stably integrated into the downstream genomic target region having the sequence of nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2, or at least 80% homology thereto, preferably into the downstream genomic target region having the sequence of nucleotides 660 to 10,260 of SEQ ID NO: 2, nucleotides 1,320 to 10,260 of SEQ ID NO: 2 or nucleotides 1,480 to 10,260 of SEQ ID NO: 2, or at least 80% homology thereto; and more preferably into the downstream genomic target region having the sequence of nucleotides 3,180 to 10,260 of SEQ ID NO: 2, nucleotides 4,920 to 9,000 of SEQ ID NO: 2 or nucleotides 6,720 to 8,460 of SEQ ID NO: 2, or at least 80% homology thereto.

[087] In another embodiment the heterologous polynucleotide stably integrated into the genome of the CHO cell of the invention or the CHO cell produced by the method of the invention is stably integrated into the upstream genomic target region and into the downstream genomic target region as disclosed above. Wherein the at least one heterologous polynucleotide integrated into the upstream genomic target region and the at least one heterologous polynucleotide stably integrated into the downstream, genomic target region may be the same or different.

[088] The skilled person will understand that a single copy, a plurality of copies of one heterologous polynucleotide, or two or more different heterologous polynucleotides may be stably integrated into the upstream genomic target region, into the downstream genomic target region, or into the upstream genomic target region and the downstream genomic target region.

[089] The at least one heterologous polynucleotide may be stably integrated into one or both alleles of the genomic target region(s).

[090] Methods for stable integration are well known in the art. Briefly, stable integration is commonly achieved by transiently introducing the at least one heterologous polynucleotide or a vector containing the at least one heterologous polynucleotide into the CHO host cell, which facilitates the stable integration of said heterologous polynucleotide(s) into the CHO cell genome. Typically the heterologous polynucleotide is flanked by homology arms, i.e., sequences homologous to the region upstream and downstream to the integration site. A vector to introduce the heterologous polynucleotide into the CHO cell of the invention may be chosen from a great variety of suitable vector systems, such as plasmids, retroviruses, cosmids, EBV-derived episomes, and the like. Various shuttle vectors may be used, e.g., vectors which may autonomously replicate in a plurality of host microorganisms such as E. coli and Pseudomonas sp. Before their introduction into the CHO host cell, circular vectors may be linearized to facilitate integration into the CHO cell genome. Methods for the introduction of vectors into CHO cells are well known in the art and include transfection with biological methods, such as viral delivery, with chemical methods, such as using cationic polymers, calcium phosphate, cationic lipids or cationic amino acids; with physical methods, such as electroporation or microinjection; or with mixed approaches, such as protoplast fusion.

[091] To enable identification or selection of recombinant cells, the at least one heterologous polynucleotide may be integrated together with a selection marker gene or a reporter gene, preferably present on the same vector. Further, the vector often includes a marker outside the homology arms allowing to identify random integration.

[092] In one embodiment the heterologous polynucleotide stably integrated into the genome of the CHO cell of the invention, or the CHO cell produced by the method of the invention are part of an expression cassette. An expression cassette comprises at least one heterologous polynucleotide coding for a gene product, such as a RNA and/or a protein, operably linked to a promoter and optionally further means controlling the expression of the gene product(s). Such means include, but are not limited to enhancers, termination signals, polyadenylation signals and a 3' untranslated region, typically containing a polyadenylation site. The promoter may be a weak promoter, or a strong promoter supporting high level expression of the gene product of interest. Said promoters include, but are not limited to CMV (cytomegalovirus) promoters, SV40 (Simian vacuolating virus 40) promoters, the RSV (Rous Sarcoma Virus) promoters, adenovirus promoters (e.g., the adenovirus major late promoter (AdMLP), CHEF-1 (CHO-derived elongation factor-1) promotors, polyoma and strong mammalian promoters such as native immunoglobulin and actin promoters or the natural promoter of the at least one heterologous polynucleotide. Preferably, the promoter is a CMV promoter or an SV40 promoter, most preferably a CMV promoter. Examples for polyadenylation

signals are BGH polyA, SV40 late or early polyA; alternatively, 3'UTRs of immunoglobulin genes etc. can be used. The skilled person will further understand that the 3' untranslated region may be engineered to support high level expression, e.g., by removing instability elements, such as AREs (adenylate-uridylate rich elements).

[093] In some embodiments, the gene product may be placed under the control of an amplifiable genetic selection marker, such as dihydrofolate reductase (DHFR), glutamine synthetase (GS). The amplifiable selection marker gene can be on the same expression vector as the secreted therapeutic protein expression cassette. Alternatively, the amplifiable selection marker gene and the secreted therapeutic protein expression cassette can be on different expression vectors, but integrate in close proximity into the host cell's genome. Two or more vectors that are co-transfected simultaneously, for example, often integrate in close proximity into the host cell's genome. Amplification of the genetic region containing the secreted therapeutic protein expression cassette is then mediated by adding the amplification agent (e.g., MTX for DHFR or MSX for GS) into the cultivation medium.

[094] Sufficiently high stable levels of the gene product in the host cell or the producer cell may be achieved, e.g., by cloning multiple copies of a heterologous polynucleotide into an expression vector. Cloning multiple copies of the heterologous polynucleotide into an expression vector and amplifying the secreted therapeutic protein expression cassette as described above may further be combined.

[095] The at least one heterologous polynucleotide encoding a gene product of interest may comprise a full length or a truncated gene, a fusion or tagged gene, and can be a cDNA, a genomic DNA, or a DNA fragment, preferably a cDNA. It can comprise the native sequence, i.e., naturally occurring form(s), or can be mutated or otherwise modified as desired. These modifications include codon optimizations to optimize codon usage in the selected host cell, humanization, fusion or tagging. The skilled person will understand that if more than one heterologous polynucleotide is stably integrated into the genome of the CHO cell of the invention or the CHO cell produced by the method of the invention, they may be encoded by more than one expression cassettes, or as part of the same expression cassette separated, e.g., by an IRES (internal ribosome entry site) sequence.

[096] In another embodiment, the heterologous polynucleotide encodes at least one protein of interest and/or at least one RNA of interest. RNAs of interest include, but are not limited to messenger RNAs (mRNAs) and small regulatory RNAs, such as microRNAs (miRNAs) or small hairpin RNAs (shRNAs). Preferably, the RNA of interest is selected from the group consisting of an mRNA, a miRNA or an shRNA, more preferably an mRNA or an shRNA. The small regulatory RNA may interfere with the expression of one or more host cell protein(s), by binding to (a) target region(s) within mRNAs coding for said host cell protein(s).

[097] The person of skill will understand that small regulatory RNAs encoded by the heterologous polynucleotide may be used to interfere with relevant processes in the host cell, such as nutrient metabolism, nutrient uptake, transcription, translation, protein folding, the unfolded protein response, apoptosis, inter- or intracellular signaling, cell cycle control, cell growth or protein secretion. Thus,

the invention can be advantageously used to engineer CHO host cells to improve their characteristics in cell culture or protein production.

[098] The RNA of interest and/or the protein of interest may be constitutively expressed or conditionally expressed. For example, expression of the RNA of interest or protein of interest may be silent during growth phase and switched on during protein production phase.

[099] The protein of interest encoded by the at least one heterologous polynucleotide stably integrated into the genome of the CHO cell of the invention or the CHO cell produced or used by the method of the invention may be a therapeutic protein selected from the group consisting of an antibody, a fusion protein, a cytokine or a growth factor, a lymphokine, an adhesion molecule, a receptor and a derivative or fragment thereof, and any other polypeptide that can serve as agonists or antagonists and/or have therapeutic or diagnostic use. Preferably the therapeutic protein is a secreted therapeutic protein. The therapeutic protein encoded by the heterologous polynucleotide may be a recombinant protein, preferably a secreted recombinant protein. Preferably, the therapeutic protein is selected from the group consisting of an antibody, a fusion protein, a cytokine or a growth factor, more preferably an antibody or a fusion protein and most preferably an antibody. Multimeric proteins, such as antibodies, may be encoded by one or more heterologous polynucleotides as part of one or more expression cassette(s).

[100] The person of skill will understand that the at least one polynucleotide stably integrated into the genome of the CHO cell of the invention or the CHO cell produced by the method of the invention may code for both, at least one RNA of interest and at least one protein of interest, advantageously combining said modification of relevant processes in the CHO cell with the expression of a heterologous protein of interest to facilitate high level and/or stable protein production, high level and/or stable protein secretion and/or a specific amount and quality of posttranslational protein modification(s).

[101] In another embodiment, the at least one heterologous polynucleotide stably integrated into the genome of the CHO cell of the invention or the CHO cell produced by the method of the invention is a marker gene. Such a marker gene may be any gene that enables a distinction between recombinant and non-recombinant cells and/or the quantification of the expression level of a gene product of interest. The marker gene may be a reporter gene or a selection marker gene. Selection markers may compensate for metabolic defects of the utilized CHO host cell, e.g. glutamine synthetase (GS) deficiency. Reporter genes may be alkaline phosphatase (AP), chloramphenicol acetyltransferase (CAT), Renilla luciferase or firefly luciferase protein(s). Reporter genes also include genes coding for fluorescent proteins, for example, green fluorescent protein (GFP) or any of the recombinant variants of GFP, including enhanced GFP (EGFP), blue fluorescent proteins (BFP and other derivatives), cyan fluorescent protein (CFP and other derivatives), yellow fluorescent protein (YFP and other derivatives) and red fluorescent protein (RFP and other derivatives). In a preferred embodiment, the reporter gene may be a fluorescent protein, such as GFP or EGFP. The selection marker may further be an antibiotic resistance gene or metabolic marker gene like aminoglycoside phosphotransferase (APH), hygromycine phosphotransferase (HYG), dihydrofolate

reductase (DHFR), thymidine kinase (TK), glutamine synthetase, asparagine synthetase and genes which confer resistance to neomycin (G418/Geneticin), puromycin, histidinol D, bleomycin, phleomycin, blasticidin and zeocin. In preferred embodiments, the selection marker gene is dihydrofolate reductase (DHFR) or glutamine synthetase (GS).

[102] In some embodiments, the at least one heterologous polynucleotide stably integrated into the genome of the CHO cell of the invention or the CHO cell produced or used by the method of the invention is part of an expression cassette. Preferably, the expression cassette is flanked by recognition sites (recognition sequence) for a site specific recombinase or a sequence specific DNA editing enzyme such as a site specific nuclease. More preferably, it is flanked by recognition sites for a site specific recombinase. Site specific recombinases are well known in the art and include, without being limited thereto, lambda integrase, PhiC31 integrase, Cre, Dre and Flp, or any derivatives thereof. Thus, the expression cassette may be flanked by recognition sites for lambda integrase, PhiC31 integrase, Cre, Dre, Flp or any derivatives thereof. Site specific nucleases include, but are not limited to zinc finger nucleases (ZFNs), meganucleases, transcription activator-like effector nucleases (TALENs) and CRISPR associated nucleases. It is well known in the art that site specific nucleases may be engineered to specifically bind a target sequence within the CHO cell genome. This facilitates the targeted exchange of DNA segments within the expression cassette enclosed by said recognition sites. The use of site specific recombinases or site specific nucleases for the targeted integration of heterologous polynucleotides into host cell genomes is routinely practiced and the respective methods are well known in the art. In some embodiments, the expression cassette comprising recognition sites for site specific recombinases or site specific nucleases may allow re-targeting of a defined genomic target region, to create multiple CHO production cells for multiple gene products, such as RNAs of interest or proteins of interest.

[103] In a specific embodiment the at least one heterologous polynucleotide stably integrated into the genome of the CHO cell of the invention or the CHO cell produced by the method of the invention is a marker gene and the marker gene is stably integrated into the CHO cell genome as part of an expression cassette and the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme (e.g., a site specific nuclease), preferably a site specific recombinase, as described above. This allows the expression cassette comprising a marker gene to be easily exchanged against an expression cassette comprising a heterologous polynucleotide coding for an RNA or a therapeutic protein of interest. Such a replacement DNA coding for a marker gene that can be easily exchanged against an expression cassette comprising heterologous polynucleotide coding a protein of interest is also referred to as “landing pad” herein.

[104] In one embodiment, the method for the production of a CHO cell according to the invention comprises the steps of (a) providing a CHO cell; (aa) introducing a first heterologous polynucleotide into said CHO cell, wherein the first heterologous polynucleotide is a marker gene and is stably integrated into the S100A gene cluster of the CHO cell genome as part of an expression cassette flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing

enzyme (e.g., a site specific nuclease), wherein (i) said heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or (ii) said heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2; and (b) introducing an expression cassette comprising a second heterologous polynucleotide into said CHO cell by replacing the expression cassette comprising the first heterologous polynucleotide of step (aa). Preferably the second heterologous polynucleotide codes for a RNA or a therapeutic protein, preferably for a therapeutic protein, more preferably a secreted protein of interest.

[105] Said first heterologous polynucleotide preferably encodes a marker gene selected from the group consisting of a reporter gene and a selection marker gene. In specific embodiments, the reporter gene may be a fluorescent protein, such as GFP. The selection marker may be dihydrofolate reductase (DHFR) or glutamine synthetase (GS). Reporter and selection marker genes may also be combined.

[106] Preferably, said first heterologous polynucleotide is integrated by targeted integration using a site-specific nuclease, more preferably by using a site-specific nuclease selected from the group of zinc finger nucleases (ZFNs), meganucleases, transcription activator-like effector nucleases (TALENs) and CRISPR associated nucleases, even more preferably by using a zinc finger nuclease (ZFN), a transcription activator-like effector nuclease (TALEN) or a CRISPR associated nuclease.

[107] Said first heterologous polynucleotide may further be part of an expression cassette flanked by recognition sites for a site-specific recombinase. Preferably, it comprises recognition sites for a site specific recombinase selected from the group consisting of lambda integrase, PhiC31 integrase, Cre, Dre and Flp.

[108] Further, an expression cassette comprising a second heterologous polynucleotide may be introduced into the CHO cell by replacing the expression cassette comprising said first heterologous polynucleotide. Preferably, said second heterologous polynucleotide encodes at least one RNA and/or at least one protein. More preferably it encodes an mRNA, miRNA or shRNA and/or a therapeutic protein. Said expression cassette comprising a second heterologous polynucleotide may be stably introduced into the CHO cell genome by targeted integration, preferably by using a site specific nuclease, or a site specific recombinase, more preferably by using a site specific recombinase, most preferably by using a site specific recombinase selected from the group consisting of lambda integrase, PhiC31 integrase, Cre, Dre and Flp.

[109] In a preferred embodiment, the method for the production of a CHO cell comprises introducing an expression cassette encompassing a first heterologous polynucleotide comprising a marker gene and recognition sites for a site-specific recombinase, wherein said first polynucleotide is stably integrated into the CHO cell genome by targeted integration, using a site specific nuclease. Further, said expression cassette encompassing the first heterologous polynucleotide is replaced by an expression cassette comprising a second heterologous polynucleotide, coding for a RNA of interest,

and/or protein of interest, by targeted integration, using a site specific recombinase. In a preferred embodiment the expression cassette comprising the first heterologous polynucleotide and the expression cassette comprising the second heterologous polynucleotide, are flanked by the same recognition site for a site specific recombinase.

[110] The person skilled in the art will understand that such a method provides a CHO cell comprising a genomic target site which is re-targetable to introduce any heterologous polynucleotides within a genomic locus supporting stable and high level expression of a gene product of interest by readily available DNA recombination methods. This may greatly reduce the time and cost associated with generating and identifying CHO production cell clones in a cell line development process.

CHO cells

[111] The CHO cell of the invention or the CHO cell produced by the method of the invention may be any Chinese hamster ovary cell capable of growing in culture and capable of expressing a RNA of interest or a protein of interest. Commonly used CHO cells for large-scale industrial production are often engineered to improve their characteristics in the production process, or to facilitate selection of recombinant cells. Such engineering includes, but is not limited to increasing apoptosis resistance, reducing autophagy, increasing cell proliferation, altered expression of cell-cycle regulating proteins, chaperone engineering, engineering of the unfolded protein response (UPR), engineering of secretion pathways and metabolic engineering.

[112] Preferably, CHO cells that allow for efficient cell line development processes are metabolically engineered, such as by glutamine synthetase (GS) knockout and/or dihydrofolate reductase (DHFR) knockout to facilitate selection with methionine sulfoximine (MSX) or methotrexate, respectively.

[113] Preferably, the CHO cell of the invention or the CHO cell produced by the method of the invention is a CHO-DG44 cell, a CHO-K1 cell, a CHO-DXB11 cell, a CHO-S cell, a CHO glutamine synthetase (GS)-deficient cell or a derivative of any of these cells.

Table 2: Exemplary CHO production cell lines

Cell line	Order Number
CHO	ECACC No. 8505302
CHO wild type	ECACC 00102307
CHO-K1	ATCC CCL-61 ECACC 85051005
CHOZN®	Merck SAFC GS -/- and DHFR -/-
CHO-DUKX (= CHO duk$^-$, CHO/dhfr$^-$,CHO-DXB11)	ATCC CRL-9096
CHO-DUKX 5A-HS-MYC	ATCC CRL-9010
CHO-DG44	Urlaub G, *et al.*, 1983. *Cell.* 33:405-412.
CHO Pro-5	ATCC CRL-1781
CHO-S	Life Technologies A1136401; CHO-S is derived from CHO variant Tobey *et al.* 1962

[114] CHO cells are most preferred, when being established, adapted, and completely cultivated under serum free conditions, and optionally in media, which are free of any protein/peptide of animal origin. Commercially available media such as Ham´s F12 (Sigma, Deisenhofen, Germany), RPMI-1640 (Sigma), Dulbecco´s Modified Eagle´s Medium (DMEM; Sigma), Minimal Essential Medium (MEM; Sigma), Iscove´s Modified Dulbecco´s Medium (IMDM; Sigma), CD-CHO (Invitrogen, Carlsbad, CA), serum-free CHO Medium (Sigma), and protein-free CHO Medium (Sigma) are exemplary appropriate nutrient solutions. Any of the media may be supplemented as necessary with a variety of compounds, non-limiting examples of which are recombinant hormones and/or other recombinant growth factors (such as insulin, transferrin, epidermal growth factor, insulin like growth factor), salts (such as sodium chloride, calcium, magnesium, phosphate), buffers (such as HEPES), nucleosides (such as adenosine, thymidine), glutamine, glucose or other equivalent energy sources, antibiotics and trace elements. Any other necessary supplements may also be included at appropriate concentrations that would be known to those skilled in the art. For the growth and selection of genetically modified cells expressing a selectable gene a suitable selection agent is added to the culture medium.

Protein production

[115] In one embodiment, the CHO cell of the invention or the CHO cell produced by the method of the invention is be used for the production of a protein of interest. The protein of interest is produced by culturing the CHO cells of the invention for a period of time sufficient to allow for expression of the antibody molecule in the host cells. Following expression, the protein of interest is harvested and may be purified. Preferably, the protein of interest is recovered from the culture medium as a secreted protein and purified using techniques well known in the art.

[116] By way of example, state-of-the art purification methods useful for obtaining the recombinant secreted therapeutic protein of the invention include, as a first step, removal of cells and/or particulate cell debris from the culture medium or lysate. The secreted therapeutic protein is then purified from contaminant soluble proteins, polypeptides and nucleic acids, for example, by fractionation on immunoaffinity or ion-exchange columns, ethanol precipitation, reverse phase HPLC, Sephadex chromatography, chromatography on silica or on a cation exchange resin. Antibodies or Fc-fusion proteins, e.g., may be purified by standard protein A chromatography, e.g., using protein A spin columns (GE Healthcare). Protein purity may be verified by reducing SDS PAGE and protein concentrations may be determined by measuring absorbance at 280 nm and utilizing the protein specific extinction coefficient. Finally, the purified recombinant secreted therapeutic protein may be dried, e.g. lyophilized.

[117] In one embodiment, the CHO cell of the invention is used to produce a protein of interest at high yield. Such production at high yield can result from high cell density, or high cell viability. It can also result from high specific cell productivity. However, the skilled person will understand that having high cell density or cell viability only supports a high total yield of the protein of interest in

case the specific cell productivity is not substantially affected or even improved. Likewise, having high specific cell productivity only supports a high total yield of the secreted recombinant therapeutic protein in case the cell density or cell viability is not substantially affected or even improved. Production at high yield thus refers to a high degree of overall productivity of the cell culture, typically measured as a concentration (titer), such as mg/mL. The production of the protein of interest according to the invention is high, if being enhanced by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 75%, at least 100% or at least 200%, compared to a control CHO cell, i.e., a CHO cell comprising the same heterologous polynucleotides randomly integrated into the genome, preferably in preferably in a CHO cell pool without clonal selection.

EXAMPLES

[118] The integration locus was identified by evaluating gene expression data. The gene S100A6, encoding a calcium binding protein, was found to be highly expressed across all experimental conditions and can therefore be used as a marker for a genomic locus supporting high heterologous protein expression. S100A6 is part of a cluster of S100 genes; hence the entire cluster was assessed.

Cell Selection and Maintenance

[119] A proprietary medium was used for the routine passaging of CHO-DG44 cells prior to complementation with an active DHFR cassette. An MTX comprising selection medium was used after transfection to enrich cells which received DHFR expression cassette integration. For CHOZN GS cell lines the commercially available medium CD Fusion (Sigma Aldrich) was used, supplemented with 850 mg/l L-Gln (6 mM). Cell lines were passaged in TPP shaker tubes at 0.3e6 cells/ml on a 2-2-3 (CHO-DG44) or 0.6e6 cells/ml on a 2-2 (CHOZN GS-/-) passaging schedule per week, respectively. Cultures were counted on an automated Vicell instrument.

[120] All of the pools went through metabolic selection by removing hypxanthine thymidine (HT) supplement from the CHO-DG44 cell line (DHFR-/-) or removing L-Gln from the CHOZN GS cell line (GS-/-). Selection was applied after the transfection process to select against parental cells that did not receive the DHFR or GS donors. Pools that have gone through the selection process were maintained in selection media permanently. When performing selection the cells were seeded at 0.4e6 cells/per ml in a T75 static flask in a total of 10-12 ml. The selection media was normally changed after 7 days by spinning down the whole flask and re-suspending the cells in the same volume of fresh media. When the selected cells “recovered” and resumed growing they were scaled up into TPP tubes as appropriate. The cells were maintained in selection media permanently prior to performance assessment in fed batch.

ZFN Technology for targeted integration

[121] The commercially available CompoZr Zinc Finger Nucleases (ZFNs) from SAFC was used for targeted integration according to assay instruction. The respective ZFNs were custom made by

SAFC based on the respective target sequence information provided, e.g., of SEQ ID NO: 11 (ZFN 13).

[122] The ZFN nucleotide sequence was unique for each ZFN arm and was linked to a FokI domain. The DNA encoding the ZNF arms was cloned into a pVAX plasmid backbone. The plasmid also contained a GFP or RFP reporter cassette upstream of the ZFN sequence, linked by a 2A peptide for separation during translation. The ZFN arms were transfected as mRNAs. For in-vitro transcription of DNA coded ZFNs into RNA the mMessage mMachine T7 Ultra kit (Ambion) was used according to the manufacturer`s instructions. Thus, of the two mRNAs used for transfection one encoded a ZNF arm specific for a sequence (e.g., ZNF13) and GFP and the other encoded a ZNF arm targeting the complementary sequence and RFP. GFP or RFP were expressed in conjunction with transfection of the ZFN arms in order to allow for quick and easy enrichment of the transfected pools by flow cytometry. Cells that have received both ZFN arms were GFP and RFP positive. These double positive cells were collected in order to create a pool enriched for ZFN activity.

Transfection Protocol

[123] For transfection a Bio-Rad Gene Pulser for electroporation was used. 1e6 cells in 2 mm cuvettes were transfected using ~20 ug of total DNA and/or mRNA (settings: 115V, 950uF, ∞ Resistance). The ZFNs are always transfected as mRNA and the donor plasmids containing the protein of interest were transfected as DNA. Cells are transfected and cultured in the same medium. Following transfection cells were cultured for 2-3 weeks to allow for washout of any transient plasmid.

Cel I Assay – ZFN Activity

[124] To measure the cleavage efficiency of ZFNs in the cell, the CEL-I or SURVEYOR nuclease assay was performed. In brief the target region was PCR amplified using genomic DNA purified from the transfected pool as the template. In the presence of active ZFNs, the genomic DNA is converted to a mixture of wild-type and NHEJ products (insertions or deletions at the target site). The PCR product was denatured under high temperatures and allowed to hybridize by gradually lowering the temperature. Some wild-type and NHEJ products hybridize to form double strand DNA with mismatches around the cleavage site, which can be cleaved by an enzyme called CEL-I or SURVEYOR resulting in cleavage products that can be separated and visualized by electrophoresis.

Junction PCR (jPCR)

[125] jPCR was used to identify sequence integration into the genome. The primers were designed to amplify the 5' or 3' ends of the donor molecule at the border of the flanking genomic DNA sequence. One primer is specific to the genomic sequence near the ZFN cut site and the second primer is specific to the donor sequence. In case the donor DNA has integrated in the correct orientation at the specified locus a PCR product is obtained. jPCR can create non-specific bands especially in pools with a combination of TI and RI events. Furthermore, the TI donors can integrate in either orientation relative to the genomic loci. Unless otherwise noted, the jPCR was performed

using primers which screen for donors that have integrated cleanly in the forward orientation. The resulting TI jPCR bands were routinely confirmed by sequencing. Parental cell line gDNA and/or donor DNA was used as negative controls.

FACS Enrichment of IgG expressing cells

[126] Flow Cytometry or Fluorescence Activated Cell Sorting (FACS) was used to enrich for certain sub-populations of cells using a FACS Aria III instrument. Typically cells were sorted for IgG expressing and GFP-negative cells, removing non-expressing cells and GFP expressing cells. Cells were prepared for FACS by spinning down and re-suspending the cells in PBS. For IgG detection cells were incubated with a fluorescently labelled anti-IgG antibody 30 min prior to sorting. A R-Phycoerythrin labelled antibody was used to bind any cells with surface bound IgG.

Productivity/Titer

[127] FACS enriched pools were assessed in a 7 or 13d fed-batch for CHO DG44 or CHOZN GS cells, respectively. The production run and titer assessment for CHO DG44 derived pools was performed with a proprietary basal medium and feed. The production CHOZN GS runs were performed in CD Fusion supplemented with Ex-Cell® CHOZN® Platform Feed. Product concentration was analysed via ForteBio Octet.

Example 1:

[128] CHO production cell clones are commonly obtained by randomly integrating heterologous polynucleotides into the host cell genome of CHO cells, i.e. by random integration (RI). Positional effects result in highly heterogeneous cell populations that consist mostly of low producer cells and only a small subpopulation of high producer cells. Additionally, high producer cells tend to be outgrown by low producer cells. To evaluate the potential of the Chinese hamster S100A gene cluster as a site for reliable, high level production of heterologous proteins (i.e. a “hot spot”), a polynucleotide encoding an IgG antibody was stably integrated into the genome of CHO-DG44 and CHOZN GS cells using a zinc finger nuclease pair engineered to be specific for a DNA sequence of SEQ ID NO: 11 (ZFN 13) as described above.

After confirming the ZFN activity and preparing donor plasmids the cells were co-transfected with the non-linearized plasmid containing the expression cassette encoding the IgG antibody and the target specific ZFN 13 pair by electroporation. Thus, the donor plasmid encoding the IgG protein of interest is being linearized randomly or via homologous recombination. Cells were cold shocked for 48 hours at 30°C to improve ZFN mRNA latency and cutting efficiency. On day four or five after electroporation, genomic DNA was harvested to perform a mismatch-specific nuclease assay, Cel I assay, to confirm ZFN activity.

[129] Following transfection the cells were cultured for 10 to 12 days before sorting to allow for complete washout of any transiently transfected donor plasmid. CHO cells were harvested by centrifugation and re-seeded in medium for metabolic selection, for CHOZN GS cells in a medium

lacking L-glutamine and for CHO-DG44 cells in a medium without hypoxanthine and thymidine supplement (HT supplement). The cultures began to recover within 5-10 days. As a control, mock cultures were transfected without plasmid and cultured in parallel. The control cultures did not exhibit growth in any experiment.

[130] Following the metabolic selection process, the cells were sorted based on GFP and IgG expression, using fluorescence-activated cell sorting (FACS) on a FACS Aria III Instrument (BD Biosciences). For IgG detection cells were incubated with a fluorescently labelled anti-IgG antibody 30 min prior to sorting. A R-Phycoerythrin labelled antibody was used to bind any cells with surface bound IgG. CHO cells were sorted into a GFP expressing population (GFP+) and a population with no GFP expression (GFP-). The donor plasmid expressing the antibody flanked by homology arms for targeted integration further contained an expression cassette encoding GFP located outside the homology arms. GFP expression was therefore associated with random integration events and the GFP negative population was enriched for cells where targeted integration occurred. The distribution and percentage of GFP+ vs GFP- cells was a good indicator for the efficiency of targeted integration and also for any positive or deleterious phenotypes at the targeted integration site. For metabolic selection, the GFP negative cell pool and the GFP positive cell pool were each cultured in 30 mL TPP tubes with a basic feed and glucose strategy. The cultures were monitored for viable cell density (VCD), viability and medium glucose levels. IgG titers in diluted supernatants were determined by direct measurement of antibody interaction using a ForteBio Octet system (Pall Biosciences) with previously established standard curves.

[131] Titers from CHO pools obtained by targeted integration (TI) or by random integration using the same polynucleotide encoding an IgG antibody for integration were measured after 3 to 7 days in batch culture for CHO-DG44 cells (Figure 1A) and after 8 to 10 days for CHOZN GS cells (Figure 1B). Titers from CHO-DG44 pools obtained by targeted integration were at least 7 fold higher than titers from CHO pool obtained by random integration titers (Figure 1A), suggesting the region upstream of the S100A3/A4/A5/A6 gene cluster is a hotspot for heterologous polynucleotide integration. Similar results were obtained for CHOZN GS cells showing at least 8 fold higher IgG titers in targeted integrated compared to random integrated cells.

Example 2:

[132] Random integration leads to cell pools that are highly heterogeneous in their expression of a heterologous protein. To evaluate if the targeted integration within the Chinese hamster S100A gene cluster leads to more homogenous expression levels and thus to a higher degree of predictability in terms of productivity, individual clones were selected from the TI cell pool and the RI cell pool of Example 1.

[133] Targeted integration and random integration pools of the CHOZN GS cells from Example 1 were used to obtain single cell clones (SCC). The process of single cloning was done by limiting dilution of the enriched TI and RI pools using conditioned medium. Conditioned medium was

prepared by culturing cells in a TPP tube at 0.3e6 cells/ml for 48 hours. Cells were sedimented and the conditioned medium was sterile filtered. The seeding was done in an 80:20 mix of cloning media (SAFC fusion platform) and conditioned media using the following steps. Step 1: Serial dilution to less than 1 cell/well were deposited in 96 well plates (200µl per well). Step 2: Cells were incubated at normal conditions and allowed to grow out for 6-7 days. Step 3: Plates were screened for single colonies of outgrowth. Wells were fed with 20µl of fresh selection medium. Step 4: Cells were cultured for about 14 days to become confluent in the 96 well plates. The cells were scaled up to a 24 well plate or harvested as needed. Step 5: gDNA for clone screening was obtained at the 96 well stage, if desired. A certain volume of cells was removed from the 96 wells and harvested using Quick Extract for subsequent PCR and sequencing. The remaining cells continued to grow out and were optionally scaled up as described in step 4. Step 6: The desired clonal populations was scaled up to TPP tubes and used for performance assessment.

[134] CHOZN GS single cell clones from random or targeted integration were assessed for protein production following cultivation for 8d in a fed-batch mode before and after 60 passages. The production runs were performed in CD Fusion supplemented with Ex-Cell® CHOZN® Platform Feed. Product concentration was analysed via ForteBio Octet and data were pooled from the same clone before and after 60 passages (n = 2 each, total n = 4).

[135] The analysis shows that single clones from populations with targeted integration exhibited highly homogeneous titers (Figure 2A) compared to single clones from populations with random integration (Figure 2B), showing that targeted integration within the S100A gene cluster resulted in predictable protein productivity. The targeted integrated clones was further more stable as reflected by the smaller error bars of the pooled data from the same clone before and after 60 passages.

Example 3:

In order to validate the hot spot locus in the S100A gene cluster, a number of additional zinc finger nucleases for TI were designed and generated as shown in Table 3 to create productive pools as described in Example 1. Figure 3A shows the location of individual ZFNs and hot spot loci in the S100A gene cluster having the NCBI Reference Sequence: NW_003613854.1. Shown are the integration sites of ZNFs 7 to 14 which are classified into “non disruptive and productive”, “non disruptive and low/non-productive” and “disruptive and low/non-productive” sites.

[136] Data was generated using CHO-ZN GS cells as described in Example 1. Eight different genomic loci were tested to evaluate whether a certain region relative to the S100A3/A4/A5/A6 main gene cluster is advantageous for the production of a heterologous gene product. It was further tested whether integration into the S100A3/A4/A5/A6 main gene cluster would lead to reduced productivity as predicted (Figure 3B).

Table 3:

Zinc finger nuclease	Targeted sequence	SEQ ID NO:

ZFN 7	tttgcttactgcccaggttctgagggaccacctggggctag	SEQ ID NO: 5
ZFN 8	cagttccctcttctgcaatattctctagctttagatgcagaa	SEQ ID NO: 6
ZFN 9	agcaactgctgtcgctcagagcttgggaggggtggatggac	SEQ ID NO: 7
ZFN 10	ccgcgcccaatgctgggaggggaagaacgggccagagcctg	SEQ ID NO: 8
ZFN 11	ctgggctgcctgcacctgtgttggctaaggctagctggttcag	SEQ ID NO: 9
ZFN 12	agcagcatctgtttccataaagtggtcaggccccaggtgggg	SEQ ID NO: 10
ZFN 13	cacaaactgaccctatgaaagtgttcagtaattcagtgccgag	SEQ ID NO: 11
ZFN 14	ggcttctactgctccagctgagcctgccctgcagtggggagg	SEQ ID NO: 12

[137] An off-target ZFN (7) integrating into the side cluster S100A1/A13/A14/A16 (comprising the nucleotide sequence of SEQ ID NO: 3) was expected to have lower expression levels, despite not interrupting any gene, due to being outside of the hotspot. Disruptive ZFNs (10, 11) integrating into the S100A3/A4/A5/A6 main gene cluster (comprising the nucleotide sequence of SEQ ID NO: 4) may damage the endogenous genes and were therefore predicted to either reduce overall achievable titers or to reduce viability. Upstream ZFNs (8 and 9) integrating into the upstream region having the nucleotide sequence of SEQ ID NO: 1 and downstream ZFNs (12, 13, 14) integrating into the downstream region having the nucleotide sequence of SEQ ID NO: 2 were expected to yield the best titers, however it was expected there may be an optimal distance from the main cluster to support protein expression.

[138] To obtain individual cell populations, CHO cells were transfected with donor plasmid and selected as described in Example 1 using the ZFNs as disclosed in Table 3. The antibody produced was the same as in Example 1. Titers of CHO pools were measured in the supernatant after 8 days of culture as described above.

[139] The actual titers resulting from targeted integration at the respective loci are shown in Fig. 3A. Off-target TI and disruptive TI (ZFNs, 7, 10, 11) did not support protein expression. Both upstream and downstream TI pools resulted in antibody titers, however, there were differences observed indicating optimal integration distances in relation to the S100A3/A4/A5/A6 main gene cluster. ZFN pair 8 supported good protein productivity, but the ZFN pair 9 site in the upstream integration region, resulted in the highest pool titers, reaching almost 0.5 g/l. The downstream ZFNs pair 13 and pair 12 both showed good protein productivity, but the more distant pair 13 relative to the S100A3/A4/A5/A6 main gene cluster showed higher titers. Further ZFN pair 14 seemed to be too far away to support adequate productivity. In conclusion, the titers showed that targeted integration disrupting genes within the S100A3/A4/A5/A6 main gene cluster or targeted outside the immediate vicinity of the S100A3/A4/A5/A6 main gene cluster resulted in low IgG production of the resulting cell populations, while integration into the region upstream and downstream of the S100A3/A4/A5/A6 main gene cluster resulted in high IgG production of the resulting cell populations. This confirms that

the S100A3/A4/A5/A6 main gene cluster is a suitable genomic target region supporting high level and reliable protein production for integration sites within genomic target regions in close distance upstream or downstream of the S100A3/A4/A5/A6 protein coding genes.

Example 4:

[140] For better applicability and easier integration of target sequences, cells may be provided comprising a "landing pad" as a replacement, such as a marker gene, at the desired location, which may be simply exchanged against the target sequence using, e.g., site directed recombination technology such as Flp-FRT recombination or Cre-lox recombination.

[141] A proprietary CHO-K1 GS cell line was used for the FRT-mediated retargeting of ZFN Locus 13 (SEQ. ID NO: 11) (landing pad approach). The respective FRT-flanked construct (see Figure 4A) was inserted using ZFN technology analogous to the method described in Example 1. Slight adaptions to meet CHO-K1 GS demands were applied to the protocol. The FRT-landing pad construct contained FRT-sites flanking a cassette containing a neomycin resistance gene, an IRES sequence and the cytosine deaminase gene (see Figure 4A). The landing pad was further flanked by an upstream and a downstream homology arm (SEQ ID NO: 13 and SEQ ID NO: 14, respectively) and the linearized construct was co-transfected together with the ZFN pair specific for locus 13 (SEQ. ID NO: 11). Correct integration was confirmed as described above and the landing pad was re-targeted (substituted) via Recombinase mediated cassette exchange (RMCE) by a gene of interest containing vector as described in the following. For routine cell culture a proprietary medium was used, supplemented with 850 mg/l L-Gln (6 mM). For maintenance of the landing pad cells 100 µg/mL G418 was used in addition.

[142] The donor sequence for exchange with the pre-integrated landing pad contained an expression cassette coding for an IgG antibody and an expression cassette coding for hygromycin. The cells stably transfected with the landing pad construct were seeded at $0.5x10^6$ cells/ml 24h prior to transfection. At the day of transfection the density of the cell culture was adjusted to $6x10^5$ cells/ml in fresh medium. 8 µg of total DNA (target vector and FLP-recombinase expressing plasmid) was diluted in CHO-S-SFMII Medium (Thermo Fisher) supplemented with L-Gln. As transfection agent PEIpro (Polyplus) was used according to the manufacturer`s manual. Following transfection the culture was kept for 24 h at 30°C and 5 % CO_2. After 24 h the temperature was switched to 36.5 °C and cultured for another 48 hours. Following transfection and selection with hygromycin only RMCE events survived. The pools were screened by junction PCR (jPCR) to confirm events in which the IgG donor has integrated into the landing pad as described above.

[143] CHO-K1 GS FRT re-targeted pools were cultured for 13 days (fed-batch) using proprietary media. Product concentrations was analysed via FortéBio Octet (Bio-Layer Interferometry (BLI) as described before. As shown in Figure 4B, IgG concentrations were increasing over time and at a very high level.

Example 5:

[144] The IgG expressing FRT targeted cells generated in Example 4 showed high homogeneity on a single clone level (Figure 5). CHO-K1 GS FRT re-targeted pools were created as described in Example 4. The process of single cell cloning was done by limiting dilution according to Example 2 with slight adaptions to the CHO-K1 GS cell line.

[145] Single-cell clones from CHO-K1 GS FRT re-targeted pools (Example 4) were cultured for 11 days in fed-batch mode using proprietary media. CHO-K1 GS cells were grown in shake flasks at 110 rpm, 36.5 °C and 5 % CO_2. The cell lines were passaged in TPP shaker tubes at $0.3x10^6$ cells/ml. Cultures are counted on automated Vi-Cell (Beckman Coulter) or Cedex Hi-Res (Roche Innovatis) instruments. As a control the respective pool was co-cultivated. Product concentration was analysed via FortéBio Octet (Bio-Layer Interferometry (BLI).

The invention is encompassed by the following items:

1. A Chinese hamster ovary (CHO) cell, comprising at least one heterologous polynucleotide, stably integrated into the S100A gene cluster of the CHO cell genome, wherein
   a) the at least one heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or
   b) the at least one heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2.

2. The CHO cell of item 1, wherein
   a) the upstream genomic target region corresponds to nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1; and/or
   b) the downstream genomic target region corresponds to nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2.

3. The CHO cell of item 1 or 2, wherein
   a) the upstream genomic target region corresponds to nucleotides 11,820 to 18,720 of SEQ ID NO: 1, nucleotides 13,560 to 18,720 of SEQ ID NO: 1, nucleotides 15,360 to 18,720 of SEQ ID NO: 1 or nucleotides 17,100 to 18,720 of SEQ ID NO: 1; and/or
   b) the downstream genomic target region corresponds to nucleotides 660 to 10,260 of SEQ ID NO: 2, nucleotides 1,320 to 10,260 of SEQ ID NO: 2 or nucleotides 1,480 to 10,260 of SEQ ID NO: 2.

4. The CHO cell of any one of items 1 to 3, wherein
   a) the upstream genomic target region corresponds to nucleotides 11,820 to 18,380 of SEQ ID NO: 1, nucleotides 13,560 to 18,380 of SEQ ID NO: 1, nucleotides 15,360 to 18,380 of SEQ ID NO: 1 or nucleotides 17,100 to 18,380 of SEQ ID NO: 1; and/or
   b) the downstream genomic target region corresponds to nucleotides 3,180 to 10,260 of SEQ ID NO: 2, nucleotides 4,920 to 9,000 of SEQ ID NO: 2 or nucleotides 6,720 to 8,460 of SEQ ID NO: 2.

5. The CHO cell of any one of the preceding items, wherein the at least one heterologous polynucleotide is stably integrated into the CHO cell genome as part of an expression cassette.

6. The CHO cell of any one of the preceding items, wherein the at least one heterologous polynucleotide codes for a RNA and/or a protein.

7. The CHO cell of item 6, wherein the RNA is a mRNA, a miRNA or a shRNA.

8. The CHO cell of item 6, wherein the at least one heterologous polynucleotide codes for a therapeutic protein, preferably a therapeutic protein selected from the group consisting of an antibody, a fusion protein, a cytokine and a growth factor.

9. The CHO cell of item 6, wherein the at least one heterologous polynucleotide is a marker gene selected from the group consisting of a reporter gene and a selection marker gene.

10. The CHO cell of item 9, wherein the marker gene is stably integrated into the CHO cell genome as part of an expression cassette and the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme.

11. The CHO cell of any one of the preceding items, wherein the CHO cell is a CHO-DG44 cell, a CHO-K1 cell, a CHO-DXB11 cell, a CHO-S cell, a CHO glutamine synthetase (GS)-deficient cell or a derivative of any of these cells.

12. The CHO cell of any of the preceding items, wherein the genomic target region consists of any one of the sequences according to claims 1 to 11 or a sequence having at least 80% sequence identity thereto.

13. The CHO cell of any one of the preceding items wherein the at least one heterologous polynucleotide is stably integrated into one or both alleles of the S100A gene cluster of the CHO cell genome.

14. A method for the production of a CHO cell, comprising the steps of

a) providing a CHO cell;

b) introducing a heterologous polynucleotide into said CHO cell, wherein the heterologous polynucleotide is stably integrated into the S100A gene cluster of the CHO cell genome, wherein

i) said heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or

ii) said heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2.

15. The method of item 14, wherein

a) the upstream genomic target region corresponds to nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1; and/or

b) the downstream genomic target region corresponds to nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2.

16. The method of item 14 or 15, wherein

a) the upstream genomic target region corresponds to nucleotides 11,820 to 18,720 of SEQ ID NO: 1, nucleotides 13,560 to 18,720 of SEQ ID NO: 1, nucleotides 15,360 to 18,720 of SEQ ID NO: 1 or nucleotides 17,100 to 18,720 of SEQ ID NO: 1; and/or

b) the downstream genomic target region corresponds to nucleotides 660 to 10,260 of SEQ ID NO: 2, nucleotides 1,320 to 10,260 of SEQ ID NO: 2 or nucleotides 1,480 to 10,260 of SEQ ID NO: 2.

17. The method of any one of items 14 to 16, wherein

a) the upstream genomic target region corresponds to nucleotides 11,820 to 18,380 of SEQ ID NO: 1, nucleotides 13,560 to 18,380 of SEQ ID NO: 1, nucleotides 15,360 to 18,380 of SEQ ID NO: 1, nucleotides 17,100 to 18,380 of SEQ ID NO: 1; and/or

b) the downstream genomic target region corresponds to nucleotides 3,180 to 10,260 of SEQ ID NO: 2, nucleotides 4,920 to 9,000 of SEQ ID NO: 2 or nucleotides 6,720 to 8,460 of SEQ ID NO: 2.

18. The method of any one of items 14 to 17, wherein the at least one heterologous polynucleotide is stably integrated into the CHO cell genome as part of an expression cassette.

19. The method of item 18, wherein the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme.

20. The method of any one of items 14 to 19, wherein the at least one heterologous polynucleotide codes for a RNA and/or a protein.

21. The method of item 20, wherein the RNA is a mRNA, a miRNA or a shRNA.

22. The method of item 20, wherein the at least one heterologous polynucleotide codes for a therapeutic protein, preferably a therapeutic protein selected from the group consisting of an antibody, a fusion protein, a cytokine and a growth factor.

23. The method of item 20, wherein the at least one heterologous polynucleotide is a marker gene selected from the group consisting of a reporter gene and a selection marker gene.

24. The method of item 23, wherein the marker gene is stably integrated into the CHO cell genome as part of an expression cassette and the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme.

25. The method of any one of items 14 to 24, wherein the heterologous polynucleotide is introduced into the CHO cell genome using

a) a sequence specific DNA editing enzyme; or
b) a site-specific recombinase.

26. The method of item 25, wherein

a) the sequence specific DNA editing enzyme is a site specific nuclease, preferably selected from the group consisting of zinc finger nucleases (ZFNs), meganucleases, transcription activator-like effector nucleases (TALENs) and CRISPR associated nucleases; and/or
b) the site specific recombinase is selected from the group consisting of lambda integrase, PhiC31 integrase, Cre, Dre and Flp.

27. The method of item 14, comprising the steps of

a) providing a CHO cell;
aa) introducing a first heterologous polynucleotide into said CHO cell, wherein the first heterologous polynucleotide is a marker gene and is stably integrated into the S100A gene cluster of the CHO cell genome as part of an expression cassette flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme, wherein
i) said heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or

ii) said heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2; and

b) introducing an expression cassette comprising a second heterologous polynucleotide into said CHO cell by replacing the expression cassette comprising the first heterologous polynucleotide of step aa).

28. The method of any one of items 14 to 27, wherein the CHO cell is a CHO-DG44 cell, a CHO-K1 cell, a CHO-DXB11 cell, a CHO-S cell, a CHO glutamine synthetase (GS)-deficient cell or a derivative of any of these cells.

29. A method for the production of a protein of interest in a CHO cell comprising

a) providing the CHO cell of any one of claims 1 to 13;

b) culturing the CHO cell of step a) in a cell culture medium at conditions allowing production of the protein of interest;

c) harvesting the protein of interest, and

d) optionally purifying the protein of interest.

30. Use of the CHO cell of any one of items 1 to 13 for producing a protein of interest at high yield.

SEQUENCE TABLE

SEQ ID NO: 1_Upstream integration locus

SEQ ID NO: 2_Downstream integration locus

SEQ ID NO: 3_Upstream side cluster

SEQ ID NO: 4_Main cluster coding area

SEQ ID NO: 5_Recognition site for ZFN 7

SEQ ID NO: 6_Recognition site for ZFN 8

SEQ ID NO: 7_Recognition site for ZFN 9

SEQ ID NO: 8_Recognition site for ZFN 10

SEQ ID NO: 9_Recognition site for ZFN 11

SEQ ID NO: 10_Recognition site for ZFN 12

SEQ ID NO: 11_Recognition site for ZFN 13

SEQ ID NO: 12_Recognition site for ZFN 14

SEQ ID NO: 13_upstream homology arm landing pad

SEQ ID NO: 14_downstream homology arm landing pad

CLAIMS

1. A Chinese hamster ovary (CHO) cell, comprising at least one heterologous polynucleotide, stably integrated into the S100A gene cluster of the CHO cell genome, wherein

   a) the at least one heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or

   b) the at least one heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2.

2. The CHO cell of claim 1, wherein

   a) the upstream genomic target region corresponds to nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1; and/or

   b) the downstream genomic target region corresponds to nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2.

3. The CHO cell of claim 1 or 2, wherein the at least one heterologous polynucleotide is stably integrated into the CHO cell genome as part of an expression cassette.

4. The CHO cell of claim 3, wherein the at least one heterologous polynucleotide is a marker gene selected from the group consisting of a reporter gene and a selection marker gene, preferably, wherein the marker gene is stably integrated into the CHO cell genome as part of an expression cassette and the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme.

5. The CHO cell of any one of the preceding claims wherein the at least one heterologous polynucleotide is stably integrated into one or both alleles of the S100A gene cluster of the CHO cell genome.

6. A method for the production of a CHO cell, comprising the steps of

   a) providing a CHO cell;

b) introducing a heterologous polynucleotide into said CHO cell, wherein the heterologous polynucleotide is stably integrated into the S100A gene cluster of the CHO cell genome, wherein

i) said heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or

ii) said heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2.

7. The method of claim 6, wherein

a) the upstream genomic target region corresponds to nucleotides 30 to 19,000 of SEQ ID NO: 1, nucleotides 2,940 to 19,000 of SEQ ID NO: 1, nucleotides 4,740 to 19,000 of SEQ ID NO: 1, nucleotides 6,480 to 19,000 of SEQ ID NO: 1, nucleotides 8,280 to 19,000 of SEQ ID NO: 1, nucleotides 10,020 to 19,000 of SEQ ID NO: 1, or nucleotides 11,820 to 19,000 of SEQ ID NO: 1; and/or

b) the downstream genomic target region corresponds to nucleotides 1 to 13,160 of SEQ ID NO: 2, nucleotides 1 to 12,000 of SEQ ID NO: 2 or nucleotides 1 to 10,260 of SEQ ID NO: 2.

8. The method of claim 6 or 7, wherein the at least one heterologous polynucleotide is stably integrated into the CHO cell genome as part of an expression cassette, preferably wherein the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme .

9. The method of any one of claims 6 to 8, wherein the at least one heterologous polynucleotide is a marker gene selected from the group consisting of a reporter gene and a selection marker gene, preferably wherein the marker gene is stably integrated into the CHO cell genome as part of an expression cassette and the expression cassette is flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme.

10. The method of any one of claims 6 to 9, wherein the heterologous polynucleotide is introduced into the CHO cell genome using

a) a sequence specific DNA editing enzyme, preferably selected from the group consisting of zinc finger nucleases (ZFNs), meganucleases, transcription activator-like effector nucleases (TALENs) and CRISPR associated nucleases; or

b) a site-specific recombinase, preferably selected from the group consisting of lambda integrase, PhiC31 integrase, Cre, Dre and Flp.

11. The CHO cell of any one of claims 1 to 3 or the method of any one of claims 6 to 8, wherein the at least one heterologous polynucleotide codes for a RNA and/or a protein.

12. The CHO cell or the method of claim 11, wherein

a) the RNA is a mRNA, a miRNA or a shRNA; and/or

b) the at least one heterologous polynucleotide codes for a therapeutic protein, preferably a therapeutic protein selected from the group consisting of an antibody, a fusion protein, a cytokine and a growth factor.

13. The CHO cell of any one of claims 1 to 5 or the method of any one of claims 6 to 10, wherein the CHO cell is a CHO-DG44 cell, a CHO-K1 cell, a CHO-DXB11 cell, a CHO-S cell, a CHO glutamine synthetase (GS)-deficient cell or a derivative of any of these cells.

14. The method of claim 6, comprising the steps of

a) providing a CHO cell;

aa) introducing a first heterologous polynucleotide into said CHO cell, wherein the first heterologous polynucleotide is a marker gene and is stably integrated into the S100A gene cluster of the CHO cell genome as part of an expression cassette flanked by recognition sites for a site specific recombinase or a sequence specific DNA editing enzyme, wherein

i) said heterologous polynucleotide is integrated upstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of SEQ ID NO: 1; and/or

ii) said heterologous polynucleotide is integrated downstream of the S100A3/A4/A5/A6 main gene cluster, into a genomic target region corresponding to the sequence of nucleotides 1 to 15,120 of SEQ ID NO: 2; and

b) introducing an expression cassette comprising a second heterologous polynucleotide into said CHO cell by replacing the expression cassette comprising the first heterologous polynucleotide of step aa).

15. A method for the production of a protein of interest in a CHO cell comprising

a) providing the CHO cell of any one of claims 1 to 5;

b) culturing the CHO cell of step a) in a cell culture medium at conditions allowing production of the protein of interest;

c) harvesting the protein of interest, and

d) optionally purifying the protein of interest.

16. Use of the CHO cell of any one of claims 1 to 5 for producing a protein of interest at high yield.