SITE-SPECIFIC INTEGRATION

Description

The present invention relates to stable and high-producing site-specific integration (SSI) host cells, e. g. Chinese hamster ovary (CHO)-derived host cells, methods to produce and to use them.

The increased numbers of biological biopharmaceutical candidates in development have fueled the need to develop robust and rapid high-throughput technologies for cell line development as the generation of commercial cell lines using conventional methods is a time-consuming, labor-intensive and repetitive process. During the construction and selection of antibody-producing cells lines, cell lines with a large range of expression, growth and stability profiles are obtained. These variations can arise due to the inherent plasticity of the mammalian genome. They can also originate from stochastic gene regulation networks or in variation in the amount of recombinant protein produced resulting from random genomic integration of a transgene principally due to the “position variegation effect”.

As a consequence of these variations and the low (1 in 10,000) frequency of genomic integration, resource-intensive and time-consuming efforts are required to screen many transfectants in the pool for these rare events, in order to isolate a commercially-compatible production cell line (e. g., a combination of good growth, high productivity and stability of production, with desired product profile). This situation can be improved by enriching for high-producing cell lines using a highly stringent selection system, such as the GS (glutamine synthase) Gene Expression System. For example, in one recent study >30% of a randomly-selected panel of 175 mAb (monoclonal antibody)-producing GS-CHOK1SV (GS Gene expression sys-

tem™, Lonza) cell lines produced >=1 g/L mAb in a fed-batch shake flask process. CHOK1SV expresses the GS enzyme endogenously, thus, positive transfectants can be obtained using glutamine-free media and selection of methionine sulfoximine (MSX). Despite such efficient selection systems, a rigorous and laborious screening regime is still required. Thus, the construction of manufacturing cell lines is a laborious and lengthy process. Not only do these cell lines need to be selected for positive growth and productivity characteristics, they also need to be cloned, and essentially need to produce the product of the correct quality for the duration of the manufacturing process. Furthermore, for a process to be economic, the cell lines generated need to exhibit consistent productivity over many cell generations. What would be desirable is to provide high-producing cell lines with positive growth characteristics with a minimum of screening activity each time a new protein of interest, e. g. a new monoclonal antibody, is to be expressed.

The technical problem underlying the present invention is to overcome the above-identified disadvantages, in particular to provide, preferably in a simple and efficient manner, high producing cell lines with a high stability and positive growth and productivity characteristics, in particular cell lines which provide a consistent productivity over a long cultivation and production period.

The present invention solves its underlying technical problem by the provision of the teaching according to the independent claims.

In particular, the present invention solves its technical problem by the provision of a site-specific integration (SSI) host cell comprising an endogenous Fer1L4 gene, wherein an exogenous nucleotide se-

quence is integrated in said Fer1L4 gene. In some embodiments, the exogenous nucleotide sequence comprises at least one gene coding sequence of interest. In some embodiments, the exogenous nucleotide sequence comprises at least two recombination target sites. In some embodiments, the recombination target sites flank at least one gene coding sequence of interest. In other embodiments, the recombination target sites are adjacent to, and do not flank, at least one gene coding sequence of interest. In some embodiments, the gene coding sequence of interest comprises at least one selection marker gene.

In a preferred embodiment the present invention foresees that the exogenous nucleotide sequences integrated in the endogenous Fer1L4 gene are two recombination target sites flanking the at least one gene coding sequence of interest, preferably the at least one selection marker gene, that means one first recombination target site being located 5' upstream and one second recombination target site being located 3' downstream to the at least one gene coding sequence of interest, preferably the at least one selection marker gene.

Preferably, the gene coding sequence of interest may be a nucleotide sequence coding for a protein of interest, e. g. an antibody, an antigen, an enzyme, a detectable protein, e. g. a fluorescent protein such as green fluorescent protein, a hormone, a growth factor, a receptor, a fusion protein, or a protein with selective function. Said nucleotide sequence may be functionally linked to at least one regulatory element, such as a promoter. Preferably, the gene coding sequence of interest is a selection marker gene.

The present invention relates in a preferred embodiment to the SSI host cell according to the present invention, wherein the recombination target site is a FRT (FLP Recognition Target) site. In a preferred embodiment of the present invention, the FRT site is a wild type FRT site, namely the F site.

In a further preferred embodiment of the present invention, the FRT site is a mutant FRT site, preferably the F5 site, preferably such as disclosed in Schlacke and Bode (1994) Biochemistry 33:12746-12752.

In a particularly preferred embodiment of the present invention the gene coding sequence of interest, e. g. the selection marker gene, is flanked at its 5' end by the wild type FRT site and at its 3' end by a mutant FRT site.

In the context of the present invention the term "recombination target sites flanking the at least one gene coding sequence of interest" means that the recombination target sites are located 5' and 3' to said gene coding sequence of interest, that means one target site is located 5' and the other target site is located 3' to the gene coding sequence of interest. The recombination target sites may be located directly adjacent or at a defined distance to the gene coding sequence of interest.

The flanking sequences, in particular the flanking recombination target sites, are positioned in forward or reverse orientation, preferably both are in forward or preferably both are in reverse orientation.

In a furthermore preferred embodiment of the present invention, the recombinant target site is a lox site.

In case the recombination target site is a FRT site, the host cells need the presence and expression of FLP (FLP recombinase) in order to achieve a cross-over or recombination event. In case the recombination target site is a lox site, the host cells needs the presence and expression of the Cre recombinase.

Both, the presence and expression of the FLP or Cre recombinase can be achieved, for example, by introduction of exogenous nucleotide sequences encoding the FLP or Cre recombinase into a host cell which nucleotide sequences are capable of being expressed in said host cell.

The present invention, thus, provides a host cell, preferably a host cell line, incorporating an exogenous nucleotide sequence, e. g. at least two recombination target sites, in particular FRT sites, and/or at least one gene coding sequence of interest, into a pre-defined "hot-spot", namely a Fer1L4 gene, that support positive combination of growth, productivity and stability. For example, in some embodiments, expression of a gene coding sequence of interest in a SSI host cell provided herein is stable over at least 70, 100, 150, 200, or 300 generations. Expression is "stable" if it decreases by less than 30%, or is maintained at the same level or an increased level, over time. In some embodiments, expression is stable if volumetric productivity decreases by less than 30%, or is maintained at the same level or is increased over time. In some embodiments, a SSI host cell provided herein produces at least 1.5 g/L, 2 g/L, 3 g/L, 4 g/L, or 5 g/L of an expression product of a gene coding sequence of interest. In some embodiments, SSI cells provided herein, in particular cell lines, are so stable they can be maintained in culture without any selection, thus present cell lines have the potential to be more acceptable to

the regulatory agencies. In terms of economics, the present invention enables rapid and resource-efficient cell line development as fewer cell lines need to be screened at different stages of the process owing to the highly predictable performance of the present cell lines. Hence, more biopharmaceutical candidates, e. g. monoclonal antibody (mAb) candidates, for a given target or more candidates for multiple targets can be developed compared to the standard process. Ultimately, this may lead to patient benefit as a result of a shortened time-to-clinic for proteins of interest, preferably therapeutic mAbs.

In the context of the present invention, the term "hot-spot" means a position, that means a site, in the genome of a host cell which provides for a stable and highly expressionally-active, preferably transcriptionally-active, production of a product, i. e. protein of a gene coding sequence of interest, in particular provides for a strong and stable production of the protein encoded by the gene coding sequence of interest, preferably wherein the gene coding sequence of interest is integrated at said position after its transfection into the host cell.

In the context of the present invention, the term "site" refers to a nucleotide sequence, in particular a defined stretch of nucleotides, i. e. a defined length of a nucleotide sequence, preferably a defined stretch of nucleotides being part of a larger stretch of nucleotides. In some embodiments, a site, e. g. a site which is a "hot-spot", is part of a genome. In some embodiments, a site is introduced into a genome, e. g. a recombination target site. A "recombination target site" is a stretch of nucleotides being necessary for and allowing, together

with a recombinase, a targeted recombination and defining the location of such a recombination.

In the context of the present invention, the term "host cell", hereinafter also called "recipient cell", refers to a cell harboring an exogenous nucleotide sequence, preferably stably integrated, in its genome.

In the context of the present invention, a "cell" is preferably a mammalian cell, in particular a rodent cell, preferably a mouse cell, a hamster cell, preferably a Chinese hamster cell, preferably a Chinese hamster ovary (CHO) cell, preferably a CHOK1 cell, preferably a CHOK1SV cell. Preferably, the cell is a human cell. Preferably, the cell is a non-human cell.

In the context of the present invention, the term "cell" preferably means cell of a cell line. Preferably, the term "cell line" refers to established immortalized cell lines.

The term "cell" in one embodiment also means primary cell.

In the context of the present invention the term "site specific integration (SSI) host cell" means a host cell comprising exogenous nucleotide sequences. In some embodiments, the exogenous nucleotide sequences include recombination target sites, enabling a site specific integration of exogenous nucleotide sequences, thus, enabling a predetermined localized and directed integration of desired nucleotide sequences at a desired place in a host cell's genome. Thus, in some embodiments, a site specific integration host cell is capable of targeted integration of gene coding sequences of interest. More preferably, a site specific integration host cell is capable of targeted integration of a gene coding sequence of interest by recombination-

mediated cassette exchange (RMCE). Preferably, such a process introduces just one functional copy of a gene coding sequence, preferably just one copy of a gene coding sequence of interest at a predetermined locus. Preferably, the process does not co-introduce vector sequences, e. g. prokaryotic vector sequences, into the host cell.

In a further preferred embodiment, two functional copies of a gene coding sequence of interest are introduced into the SSI host cell.

In the context of the present invention, the term "selection marker gene" refers to a nucleotide sequence, in particular a gene coding sequence, that means a protein-coding nucleotide sequence, hereinafter also called region, under regulatory and functional control at least one regulatory element, in particular a promoter, wherein said protein-coding region encodes a protein allowing for selection of host cells expressing said protein.

In the context of the present invention, the term FRT means FLP Recognition Target. The FRT is a 34 base pair long nucleotide sequence which enables a site-directed recombination technology allowing the manipulation of an organism DNA under controlled conditions in vivo. The FRT is bound by the FLP recombinase (FLP) which subsequently cleaves said sequence and allows the recombination of nucleotide sequences integrated between two FRT sites. For RMCE, two cross-over events are required mediated by two flanking recombinase target sequences; one at the 5' and one at the 3' end of the cassette to be exchanged. A cross-over can occur between two identical FRT sites. The use of FRT sites also requires the expression and presence of the FLP recombinase. The whole system, herein also called "FRT/FLP", is disclosed in Seibler and Bode, Bio-

chemistry 36 (1997), pages 1740 to 1747, and Seibler et al., Biochemistry 37 (1998), pages 6229 to 6234.

In the context of the present invention, a Fer1L4 gene is the Fer1L4 wild type gene, all of its isoforms and all of its homologues, in particular as long as the homologues have a sequence homology of at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 95, at least 96, at least 97, at least 98, at least 99 or at least 99.5 % to the wild type Fer1L4 gene, preferably over the full length of the wild type Fer1L4 gene, preferably the wild type hamster Fer1L4 gene, preferably having the coordinates 1176191 to 1781992 from NCBI accession number NW_003613833 or JH000254.1, preferably the wild type CHO Fer1L4 gene or a shortened form thereof.

Most preferably, the wild type Fer1L4 gene present in the present SSI host cells is characterized by firstly, the 5' integration site of the 5' located flanking sequence is located between exon 39 and 40 and secondly, the 3' integration site of the 3' located flanking sequence is located between exon 28 and 29. Thus, in one preferred embodiment, the integration of the exogenous sequences involves parts of the endogenous Fer1L4 gene, preferably the region spanning and including exons 28 to 40. In some embodiments, at least a portion of the Fer1L4 gene is deleted in an SSI host cell.

In the present invention, "homologues" or "homologous sequences" are nucleotide sequences, which have the above identified sequence homology to the specifically given comparative sequence, e. g. to the wild type CHO Fer1L4 gene or parts thereof, e. g. SEQ ID No. 7, 8 or 9.

In the context of the present invention, the term "sequence homology" refers to a measure of the degree of identity or similarity of two sequences based upon an alignment of the sequences which maximizes similarity between aligned nucleotides, and which is a function of the number of identical nucleotides, the number of total nucleotides, and the presence and length of gaps in the sequence alignment. A variety of algorithms and computer programs are available for determining sequence similarity using standard parameters. Preferably, sequence homology is measured using the BLASTn program for nucleic acid sequences, which is available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/), and is described in, for example, Altschul et al. (1990), J Mol. Biol. 215:403 -410; Gish and States (1993), Nature Genet. 3:266-272; Madden et al. (1996), Meth. Enzymol. 266: 131-141; Altschul et al. (1997), Nucleic Acids Res. 25:33 89-3402); Zhang et al. (2000), J. Comput. Biol. 7(l-2):203-14. Preferably, sequence homology of two nucleotide sequences is the score based upon the following parameters for the BLASTn algorithm: word size = 11; gap opening penalty = -5; gap extension penalty = -2; match reward = 1; and mismatch penalty = -3.

The Fer1L4 gene may thus be the CHO Fer1L4 gene itself or may also be e. g. a human Fer1L4 gene, for instance Dysferlin (Fer1L1), Otoferlin (Fer1L2), Myoferlin (Fer1L3), Fer1L4, Fer1L5, or Fer1L6.

Preferably, the present Fer1L4 gene may also be the C. elegans Fer1 gene (NCBI Gene ID: 172659, WormBase: WBGene00001414) or any other member of the ferlin family of which there are six members in mammalian cells. Ferlins facilitate vascular fusion, specifically membrane fusion events. C. elegans Fer1 is required for the fusion

of organelles to plasma membrane and normal reproduction in worms (Achanzar, W. E., and Ward, S., J. Cell Sci. 110 (1997), 1073–108; Washington, N. L., and Ward, S., J. Cell Sci. 119 (2006), 2552–2562).

Along with a C-terminal anchor, the mammalian ferlin family members also contain multiple (6 or 7) C2 domains (Sutton RB et al. Cell 80 (1995), 929-38; Shao et al. Science 273 (1996), 248-251). Thus, genes encoding C2 domains are hereinafter also considered to be homologous to the present wild type Fer1L4 gene.

In the context of the present invention, the term “exogenous gene” or “exogenous nucleotide sequence” refers to a nucleotide sequence introduced into a host cell, e. g. by conventional genetic engineering methods, preferably by means of transformation, electroporation or transfection, which was prior to said introduction not present in said host cell. Such sequences are also termed “transgenic”.

The term “endogenous gene” or “endogenous nucleotide sequence” refers to a nucleotide sequence originating from and being present in a host cell and therefore is not being introduced therein from outside said host cell.

The term “nucleotide sequence” or “polynucleotide” as used herein refers preferably to nucleic acids, preferably a DNA or RNA.

In a preferred embodiment of the present invention, a gene coding sequence of interest flanked by the at least two recombination target sites is a selection marker gene or a gene coding sequence encoding an antibody, e. g. a monoclonal antibody, an antibody derivative, a fusion protein, an enzyme or a biologically active protein, e. g. a

growth factor or peptide hormone, G-CSF, GM-CSF, EPO, TPO, an interleukin, an interferon etc., in particular a pharmaceutically or nutritionally functional protein. Preferably, the gene coding sequence of interest is exogenous to the host cell.

In furthermore preferred embodiments, the gene coding sequence of interest may also be a structurally or functionally defined part of a gene, for instance a fragment of an antibody, such as a heavy or light chain thereof or a part of a functional protein. In some embodiments, a gene coding sequence of interest may encode an expression product comprising a structurally or functionally defined part of a polypeptide, e.g., a discrete domain, set of domains, or portion of a domain, such as a heavy or light chain of an antibody or a constant region of an antibody.

The present invention relates in a preferred embodiment to the SSI host cell according to the present invention, wherein the selection marker is the GS selection marker, the hygromycin selection marker, the puromycin selection marker or the thymidine kinase selection marker.

In the context of the present invention, the GS selection marker, encoded by a GS marker gene, operates in a GS marker system. Accordingly, in the absence of glutamine in the growth medium, the glutamine synthetase (GS) activity is essential for the survival of mammalian cells in culture. Some mammalian cell lines, such as mouse myeloma lines, do not express sufficient GS to survive without added glutamine. With these cell lines a transfected GS marker gene can function as a selectable marker by permitting growth in a glutamine-free medium. Other cell lines, such as Chinese hamster

ovary cell lines, express sufficient GS to survive without exogenous glutamine. In these cases, the GS inhibitor methionine sulfoximine (MSX) can be used to inhibit endogenous GS activity such that only transfectants with additional GS activity can survive.

The present invention relates in a preferred embodiment to the SSI host cell according to the present invention, wherein the host cell is a CHO host cell or a CHOK1SV (Porter, AJ et al. Biotechnol Prog. 26 (2010), 1455-1464) host cell.

The present invention also relates in a preferred embodiment to SSI host cells, wherein the exogenous sequences are integrated at a location spanning from position 1750049 (5' integration site) to 1760965 (3' integration site) (see figure 7). The flanking sequences are preferably located at scaffold coordinates 1750049 to 1750870 (5' end, SEQ ID No.9, 822 bp) and 1758964 to 1760965 (3' end, SEQ ID No. 7 and 8, 2000bp) (see figure 7).

The present invention relates in a preferred embodiment to the SSI host cell of the present invention, wherein the nucleotide sequences of the Fer1L4 gene flanking the integrated exogenous nucleotide sequences, namely the at least one gene coding sequence of interest which itself is flanked at its 5' and 3' end by one recombination target site each, are selected from the group consisting of SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 9 and a homologous sequence thereof.

In a preferred embodiment of the present invention, the flanking sequences being homologous to the sequences given in SEQ ID No. 7, 8 or 9 have a sequence homology of at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 95, at least 96, at

least 97, at least 98, at least 99 or at least 99.5 % to these regions of the wild type Fer1L4 gene, preferably over their full length.

Thus, in a particularly preferred embodiment, the SSI host cell of the present invention is characterized by the presence of exogenous nucleotide sequences, namely the at least one gene coding sequence of interest, which itself is flanked at its 5' and 3' end by one recombination target site each, and wherein at least one of the nucleotide sequences of SEQ ID No. 7 or 8 or a homologous sequence thereof is located at the 3' end of the exogenous nucleotide sequences integrated into the genome of the host cell.

Thus, in a particularly preferred embodiment the SSI host cell of the present invention is characterized by the presence of exogenous nucleotide sequences, namely the at least one gene coding sequence of interest, which itself is flanked at its 5' and 3' end by one recombination target site each, and wherein at least one nucleotide sequence as given in SEQ ID No. 9 or a homologous sequence thereof is located at the 5' end of the exogenous nucleotide sequences integrated into the genome of the host cell.

Preferably, the SSI host cell comprises the exogenous nucleotide sequence, namely at least one gene coding sequence of interest which itself is flanked at the 5' and 3' end by one recombination target site each, which is flanked at its 3' end by at least one of the nucleotide sequences of SEQ ID No. 7, 8 or a homologous sequence thereof and at its 5' end by a nucleotide sequences given in SEQ ID No. 9 or a homologous sequence thereof.

In a particularly preferred embodiment, the given 5'- and/or 3'-flanking sequences of SEQ ID No. 7, 8, 9 or a homologous sequence

thereof are located directly adjacent and without any intervening sequences to the 5' end or to the 3' end or to the 5' end and 3' end of the recombination target site(s) integrated in the Fer1L4 gene.

In a furthermore preferred embodiment of the present invention there is provided an isolated nucleotide molecule, preferably polynucleotide, comprising a portion of a Fer1L4 gene, e. g. comprising at least one nucleotide sequence selected from the group consisting of the nucleotide sequences as given in SEQ ID No. 7, 8, 9 and a homologous sequence thereof.

Particularly preferred is an isolated nucleotide molecule, preferably polynucleotide, consisting of at least one nucleotide sequence selected from the group consisting of the nucleotide sequences as given in SEQ ID No. 7, 8, 9 and a homologous sequence thereof.

In a furthermore preferred embodiment of the present invention, a vector, preferably an expression vector, comprising the isolated nucleotide molecule of the present invention, in particular SEQ ID No. 7, 8, 9 or homologues thereof and a transfected host cell comprising said vector or said nucleotide molecule is provided.

The nucleic acid sequences defining the Fer1L4 locus, i. e. the nucleic acid sequence of the Fer1L4 gene, i. e. the Fer1L4 nucleotide sequence, in particular nucleotide sequences selected from the group consisting of nucleotide sequences as given in SEQ ID No. 7, 8, 9 and homologous sequences thereof, were herein empirically identified by sequences upstream and downstream of the integration site of a nucleic acid construct comprising an expression cassette of a cell line expressing a reporter protein at a high level. These nucleic acid sequences of the invention provide sequences with a new func-

tionality associated with enhanced and stable expression of a nucleic acid, for example, an exogenous nucleic acid comprising a gene coding sequence of interest that appear to function differently from that previously described for cis-acting elements such as promoters, enhancers, locus control regions, scaffold attachment regions or matrix attachment regions.

The present nucleotide sequences do not appear to have any open reading frames (ORFs), making it unlikely that they encode transactivator proteins. Transfection experiments demonstrated that the present sequences display some characteristics of cis-acting elements. The present sequence activity is not detected in transient transfection assays; the present sequences also appear to be distinct from promoter and enhancer elements, which are detected with these methods.

The present invention also relates to the use of a Fer1L4 nucleotide sequence, in particular a nucleotide sequence selected from the group consisting of nucleotide sequences as given in SEQ ID No. 7, 8, 9 and homologous sequences thereof, in a vector, in particular an expression vector, in particular a non-RMCE expression vector comprising at least one gene coding sequence of interest, in particular for producing cell lines, preferably in a random process, in particular for producing cell lines showing enhanced expression, in particular for producing high producer cell lines, preferably in higher frequencies and which preferably provide greater productivity stability. Preferably, such a use foresees the transfection of cells with the above-identified Fer1L4 nucleotide sequences, preferably vectors containing said nucleotide sequences and obtaining stably transfected cell lines therefrom.

Thus, the present invention also relates to a method of producing cells or cell lines, preferably high producer cells or cell lines with high productivity stability, wherein Fer1L4 nucleotide sequences, in particular nucleotide sequences selected from the group consisting of nucleotide sequences as given in SEQ ID No. 7, 8, 9 and homologous sequences thereof, preferably integrated in a vector, preferably an expression vector, preferably a non-RMCE expression vector, are transfected into cells or cell lines and the stably transfected cells or cell lines are selected and obtained. The presence of sequences of the Fer1L4 locus or parts thereof as identified herein provides a cis-acting effect to genes of interest thereby enhancing their expression wherever they are integrated in the genome. Cell lines generated this way in a random process are expected to show greater productivity stability.

The present invention preferably relates to a use of a Fer1L4 nucleotide sequence in an expression vector for the production of a stable and highly transcriptional active cell line or cell.

The present invention preferably relates to the above-identified use, wherein the Fer1L4 nucleotide sequence is selected from the group consisting of nucleotide sequences as given in SEQ ID No. 7, 8, 9 and homologous sequences thereof.

The present invention preferably relates to a method for the production of a product of a gene coding sequence of interest comprising cultivating a host cell produced according to the present method for producing a cell or cell line in a suitable medium and recovering the product therefrom.

The present invention relates in a preferred embodiment to an SSI host cell according to the present invention, wherein the exogenous nucleotide sequences include at least one wild type FRT site, preferably at least two wild type FRT sites, flanking at least one gene coding sequence of interest, or at least one mutant FRT site, preferably at least two mutant FRT sites flanking at least one gene coding sequence of interest. Most preferably, the exogenous nucleotide sequences are one wild type FRT site and one mutant FRT site flanking at least one gene coding sequence of interest. Particularly preferred, a gene coding sequence of interest, preferably a selection marker gene, is located between a wild type FRT, preferably located 5' to the gene coding sequence of interest, preferably the selection marker gene, and a mutant FRT site, preferably located 3' to the gene coding sequence of interest, preferably the selection marker gene. This preferably ensures that recombination-mediated cassette exchange always occurs in the same orientation.

The present invention relates in a preferred embodiment to a method for producing a SSI host cell according to the present invention comprising the steps of a) transfecting a cell, preferably comprising an endogenous Fer1L4 gene, with a first vector comprising a first exchangeable cassette, the cassette comprising at least two recombination target sites, in particular FRT sites, flanking at least one first gene coding sequence of interest, preferably coding for a mAb, subsequently b) selecting transfected cells comprising the at least two recombination target sites, in particular FRT sites, flanking at least one first gene coding sequence of interest integrated in the endogenous Fer1L4 gene and showing a high and stable production of the product of the first gene coding sequence of interest; subsequently c) transfecting the cells obtained in step b) with a second vector com-

prising a second exchangeable cassette, the cassette comprising at least two matching recombination target sites, in particular FRT sites, flanking at least one second gene coding sequence of interest, namely a selection marker gene, subsequently d) effecting a site-directed recombination-mediated cassette exchange and subsequently e) selecting transfected cells expressing the second gene coding sequence of interest, preferably the selection marker gene, so as to obtain the SSI host cell comprising the second exchangeable cassette stably integrated in its genome according to the present invention (see for instance figure 4).

Preferably, the first and second gene coding sequence of interest is flanked at one of its ends by a first recombination target site and at its other end by a second recombination target site which is different to the first target site.

Thus, the present invention provides for a method for producing a SSI host cell according to the present invention which method uses a recombinase-mediated cassette exchange and in the course of which a first exchangeable cassette comprising a first gene coding sequence of interest, preferably coding for a mAb, flanked by recombination target sites is transfected via a vector into a host cell, integrated into the host cell's genome, in particular a "hot-spot" thereof, and after selection of "hot-spot" transfectants a second exchangeable cassette, for instance being part of the circular exchange plasmid and being composed of at least one gene coding sequence of interest, in particular a second gene coding sequence of interest, preferably a selection marker gene, being flanked by matching recombination target sites is transfected into the host cell and allowed to recombine thereby exchanging the first gene coding sequence of

interest by the second gene coding sequence of interest (see for instance figure 4). Advantageously, only one copy of the gene coding sequence of interest is inserted at the predetermined locus and no vector sequences, in particular plasmid sequences of the exchange plasmid are integrated into the host genome.

The present method for producing an SSI host cell is advantageous in so far as in steps a) and b) a "hot-spot" showing a high and stable production of a product of a gene coding sequence of interest can be identified using at least a first gene coding sequence of interest, for instance a gene encoding an antibody, e. g. a mAb, or a part thereof and being a gene of industrial utility, for instance a biopharmaceutically relevant protein, and that in steps c), d) and e) said first gene coding sequence of interest being used to identify the interesting "hot-spot" is completely exchanged by a so-called null cassette, in particular by a second exchangeable cassette comprising at least one second gene coding sequence of interest, namely one selection marker gene. In this way an SSI host cell free of pre-existing sequences of the first gene coding sequence of interest used for the identification of the "hot-spot" is created which allows a further recombinase-mediated cassette exchange to place at said "hot-spot" another gene coding sequence of interest, namely a third gene coding sequence of interest replacing the second gene coding sequence of interest, preferably the selection marker gene. Thus, the presently obtained SSI host cells can be used to effect a further recombinase-mediated cassette exchange to place a third gene coding sequence of interest into the genome of the host cell at the identified "hot-spot".

In the context of the present invention, the term "matching recombination target sites" means that a first site of said recombination tar-

get sites of an exchangeable cassette is identical to a first recombination target site of another exchangeable cassette and that a second recombination target site of the firstly mentioned exchangeable cassette is identical with the second recombination target site of the other exchangeable cassette thereby allowing an exchange of the nucleotide sequences in between the recombination target sites. Preferably, the first recombination target site of both exchangeable cassettes is different to the second recombination target site of both exchangeable cassettes.

The present invention relates in a preferred embodiment to a method for producing the product of a gene coding sequence of interest comprising the steps of i) transfecting an SSI host cell according to the present invention with a vector comprising at least one third exchangeable cassette, which cassette comprises at least two matching recombination target sites flanking at least one third gene coding sequence of interest and at least one selection marker, ii) effecting a site-directed recombination-mediated cassette exchange so as to obtain an SSI host cell comprising the third gene coding sequence of interest, iii) allowing the SSI host cell obtained in step ii) to express the third gene coding sequence of interest and iv) recovering the product of the third gene coding sequence of interest.

One major advantage of the present invention is the provision of an intrinsic production stability which is inherited from the SSI host cell generated in step b) comprising the first gene coding sequence of interest in an identified “hot-spot”, namely the Fer1L4 gene, all the way through to an SSI host cell comprising the third gene coding sequence of interest. This stability is independent of selection and allows the conclusion that the “hot-spot” identified in steps a) and b) as

features associated with sequences around the "hot-spot" or elsewhere in the genome. In a cell line construction based on random vector integration this is a very rare event and the effort prior to find such a site is immense. Thus, the present invention allows the elimination of elongated stability studies for the selection of suitable cell lines resulting in an overall shorter developmental cycle and resource reduction. Furthermore, the present invention allows to culture cells without selection after the recombination-mediated cassette exchange, meaning that the manufacture process is potentially more acceptable to regulatory agencies. Advantageously, the present "Fer1L4 hot spot" was defined and identified using a mAb, preferably not using a fluorescence marker.

The present invention provides for a method of producing an SSI host cell in which an exogenous nucleotide sequence is introduced into an endogenous Fer1L4 gene in a cell by targeted integration. Any of a variety of methods for directing a nucleotide sequence into a specified site of interest in the genome can be employed, and include homologous recombination, and nuclease mediated methods e. g. use of parvovirus-mediated homologous recombination, use of zinc finger nucleases, transcription activator-like effector nucleases, or engineered meganucleases (see, e. g. Russell and Hirata, Nat. Med. 18(4):325-30, 1998; US Pat. Pub. No. 20120070890; US Pat. No. 6,528,313; US Pat. Pub. No. 20090258363). An exogenous nucleotide sequence introduced by such a method can include any of the features described herein. For example, an endogenous nucleotide sequence can include at least one gene coding sequence of interest and/or at least two recombination target sites. In some embodiments, the gene coding sequence of interest comprises at least one selection marker gene.

In a further embodiment the present invention relates to a method of producing an SSI host cell, the method comprising introducing an exogenous nucleotide sequence into an endogenous Fer1L4 gene in a cell. Preferably, the exogenous nucleotide sequence is introduced by homologous recombination between the Fer1L4 gene and a polynucleotide, wherein the polynucleotide comprises a) a first nucleotide sequence homologous to a first portion of the Fer1L4 gene, b) the exogenous nucleotide sequence, and c) a second nucleotide sequence homologous to a second portion of the Fer1L4 gene. In this further embodiment, the introduction of the exogenous nucleotide sequence is preferably facilitated using a viral vector e.g., an adeno-associated virus vector which mediates homologous recombination or an exogenous nuclease e.g., a zinc finger nuclease, a transcription activator-like effector nuclease, or an engineered meganuclease. Particularly preferred, an adeno-associated virus vector is used. In a particularly preferred embodiment, the exogenous nucleotide sequence is flanked by recombination target sites, preferably loxP sites.

The present invention relates in an aspect A also to a method for producing an SSI host cell comprising the steps of a) providing a cell comprising an endogenous Fer1L4 gene, wherein an endogenous nucleotide sequence is integrated in said Fer1L4 gene, and wherein the endogenous nucleotide sequence comprises at least two recombination target sites, flanking at least one first gene coding sequence of interest; subsequently b) transfecting the cells provided in step a) with a vector comprising a first exchangeable cassette, the cassette comprising at least two matching recombinant target sites, flanking at least one second gene coding sequence of interest, namely a selection marker gene, subsequently c) effecting a site-directed recombi-

nation-mediated cassette exchange and subsequently d) selecting transfected cells expressing second gene coding sequence of interest, preferably the selection marker gene, so as to obtain the SSI host cell comprising the first exchangeable cassette stably integrated in its genome.

The present invention relates also to a method for producing an SSI host cell comprising the steps of a) providing a cell comprising an endogenous Fer1L4 gene, wherein an exogenous nucleotide sequence is integrated in said Fer1L4 gene, and wherein the exogenous nucleotide sequence comprises at least two recombination target sites, flanking at least one first gene coding sequence of interest; subsequently b) transfecting the cells provided in step a) with a vector comprising a first exchangeable cassette, the cassette comprising at least two matching recombinant target sites, flanking at least one second gene coding sequence of interest, namely a selection marker gene, subsequently c) effecting a site-directed recombination-mediated cassette exchange and subsequently d) selecting transfected cells expressing second gene coding sequence of interest, preferably the selection marker gene, so as to obtain the SSI host cell comprising the first exchangeable cassette stably integrated in its genome.

The present invention also relates to the above-identified method of aspect A further comprising i) transfecting the SSI host cell with a vector comprising at least one second exchangeable cassette, which cassette comprises at least two matching recombination target sites flanking at least one third gene coding sequence of interest and at least one selection marker, ii) effecting a site-directed recombination-mediated cassette exchange so as to obtain a SSI host cell compris-

ing the third gene coding sequence of interest, iii) allowing the SSI host cell obtained in step ii) to express the third gene coding sequence of interest and iv) recovering the product of the third gene coding sequence of interest.

The present invention also provides the use of a cell, preferably a CHO cell, comprising an endogenous Fer1L4 gene, preferably a CHO Fer1L4 gene, for the production of a stable and highly transcriptionally-active cell line.

Before the present invention is described in more detail, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, some potential and preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. It is understood that the present disclosure supersedes any disclosure of an incorporated publication to the extent there is a contradiction.

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

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

Further preferred embodiments are the subject matter of the subclaims.

SEQ ID No. 1 to 3 represent nucleotide sequences of vectors used in the present invention,

SEQ ID No. 4 to 6 represent primers used in the present invention,

SEQ ID No. 7 and 8 represent the 3'-located sequences of the Fer1L4 CHO gene,

SEQ ID No. 9 represents a 5'-located sequence of the Fer1L4 CHO gene and

SEQ ID No. 10 and 11 represent further primers used in the present invention.

The invention will be further described by way of examples and the accompanying figures.

The figures show:

Figure 1 shows the vector pRY17 (SEQ ID No. 1) used to identify the "hot-spot" in CHOK1SV. LC1 and HC1 are the light and heavy chains of chimeric monoclonal IgG4 antibody, cB72.3. hCMV refers to the hCMV-MIE early gene promoter where 'Int' denotes its first intron, Intron A and the flanking exons encoding the 5' UTR are denoted as 'Ex1' and 'Ex2' respectively. The wild type and F5 mutant FRT sites are labeled, F and F5, respectively. Polyadenylation, SV40 early promoter and β-lactamase sequences are indicated as pA, SV40 and bla, respectively. GS denotes the glutamine synthetase cDNA and Hyg (-ATG) denotes a hygromycin phosphotransferase gene lacking an initiation methionine codon and promoter.

Figure 2 shows the 'Intermediate' vector, pRY37 (SEQ ID No.2) used for to create SSI host 10E9. This vector contains a mutant FRT (F5) and wild type FRT site (F) flanking transcription units containing thymidine kinase (TK) and puromycin acetyl transferase (PAC) genes. Transcription in each is driven by the SV40 early promoter (SV40E). Vector pRY37 was co-transfected into 11A7 cells with vector pOG44 (Invitrogen) that encodes the FLP recombinase. Cell lines were selected in the presence of puromycin-containing medium and further screened for RMCE between the pRY37 FRT sequences and the equivalent sequences of pRY17 embedded in the 11A7 genome (see figure 3). After screening and cloning, cell line 10E9 was isolated (the 'SSI host') in which the cB72.3 mAb transcription units had been exchanged for those of TK and PAC.

Figure 3 shows the targeting vector for creating mAb-producing cell lines in SSI host 10E9. This vector, pRY21 (SEQ ID No. 3) contains transcription units containing the Myo mAb genes (HC2 and LC2), flanked by mutant (F5) and wild type FRT sites. An in-frame initiation methionine codon has been added to the 5' end of the wild type FRT site (ATG+F) and upstream of this is an SV40 early promoter (SV). Transcription of the HC2 and LC2 genes is driven by the promoter of the hCMV major immediate early gene 1 (hCMV-MIE) and its first intron, Intron A (Int A) and the flanking exons encoding the 5' UTR denoted as 'Ex1' and 'Ex2', respectively. The β-lactamase gene is denoted as bla.

Figure 4 shows the process of generating SSI host 10E9 from the 11A7 cell line. This schematic diagram shows a single copy of linearly genomically integrated pRY17 containing mAb cB72.3 transcription units (HC1 and LC1) flanked by a mutant F5 and wild type FRT site in cell line 11A7. Prior to transfection, vector pRY17 was cut in the middle of the β-lactamase gene (bla) with Pvu I, the linear vector is then flanked by 5' (bla L) and 3' (bla R) portions of the gene. Vector pRY37 is co-transfected into cell line 11A7 with a plasmid encoding FLP recombinase and recombination (indicated by crosses) occurs between similar sites leading to RMCE of the mAb transcription units with thymidine kinase (TK) and puromycin acetyl transferase (PAC) transcription units each driven by the SV40 early promoter (SV).

Figure 5 shows the process of generating a mAb-producing cell line from the SSI host, 10E9. The targeting vector, pRY21 (figure 3, SEQ ID No. 3) contains transcription units for the second (Myo) mAb (HC2 and LC2) flanked upstream by a mutant FLP site (F5) and down-

stream by an SV40 early promoter (SV) itself upstream of wild-type FRT site linked in-frame with methionine initiation codon (ATG-F). In the presence of Flp recombinase, and the targeting vector pRY21 the thymidine kinase (TK) and puromycin acetyl transferase (PAC) genes and associated promoters are substituted by RMCE with the new Myo mAb antibody transcription units. As a consequence of this, a functional fusion gene of the wild type FRT site and hygromycin gene is created (ATG-F-Lnk-Hyg). Therefore cells where specific RMCE had occurred can be selected in the presence of hygromycin and ganciclovir.

Figure 6 shows a comparison of antibody expression of SSI pools and of clonal cell lines derived from random process or FLP-aided site-specific integration (SSI). Random Integration clones were generated using a process identical to the one used to generate cell line 11A7 in Phase 1. SSI pools (SSI Round 1) and clonal cell lines (SSI Round 2, and SSI Round 3) were generated under different transfection and selection conditions as specified in the method section. For SSI Round 3, 4 different selection conditions were employed; 200 µg/mL, without MSX (●), 200 µg/mL hygromycin with MSX (◆), 400 µg/mL hygromycin without MSX (▲) or 400 µg/mL hygromycin with MSX (▼). Clonal cell lines generated from all process were expanded into shake flasks. The productivity of clonal cell lines was assessed in 7 day terminal batch shake-flasks (see methods).

Figure 7 shows the exon structure of the Fer1L4 gene on the minus strand of scaffold1492 aka. JH000254.1 (http://www.ncbi.nlm.nih.gov/nuccore/JH000254). The location of the 5' and 3' integration sites have been indicated (1750049 and 1760965, respectively).

Figure 8 shows the mapping of 10E9 reads mapped to pRY17-GA-Q using BWA. Though inspection of the mapping, it is found that multiple unpaired reads (black arrows) mapped at 5' end of HC fragment (which has 214 bp deleted from 5' end of the complete HC).

Figure 9 shows the structure of the landing pad in the Fer1L4 gene of the SSI host cell 10E9. (**A**) The WGRS coverage for the landing pad is shown superimposed over the top of the 'single-copy' model and below a more detailed schematic is shown. (**B**) The Southern blot data suggested that there may be two copies of the landing pad. Genomic DNA from 10E9 cells was digested with either Bmg I or Pml I restriction enzymes and analysed by Southern Blot. The Southern blots were hybridised with a TK probe, revealing two fragments that span tow copies of the landing pad (shown mapped to a 'two-copy' model). The 'two-copy' model is consistent with the WGRS data.

Figure 10 shows the schematic drawing of integration sites of 10E9 (SSI host cell) and RMCE cell lines (expressing a gene of interest). The genomic location is denoted by a dashed vertical line and nucleotide coordinate on the unplaced CHO-K1 Scaffold1492 scaffold (accession number JH000254.1, identical to NW_003613833.1). The Fer1L4 gene is depicted as an arrow whilst the genomic fragment is represented as a solid line. The interrupting dashed line represents a very large distance. 5' and 3' flanking sequences are represented by white and black pointers, respectively.

Figure 11 shows coverage graph of mapped whole genome resequencing reads on the one copy model of integrated vector. The height of the grey chart is equal to the coverage at each base without smoothing (shown for each RMCE-generate cell line, 1-4). The cov-

erage at the flanking region is about the same among all cell lines. However the coverage of the two high producers on the top of the figure is about 1.4 to 2.3 times of the coverage of the two low producers in the bottom. The wFRT features are indicated by asterisks, whilst the mFRT features are depicted by black circles. (Other features are indicated as follows: HC with 214bp deletion, is the remaining cB72.3 left behind in the hot-spot after the creation of the SSI host 10E9; BlaR is a 5' portion of the beta-lactamase gene; Myo-LC and Myo-HC are the anti-Myostatin mAb HC (heavy chain) and LC (light chain) encoding sequences, resident in the hot-spot post-RMCE; Hyg = hygromycin; BlaL is the 3' portion of the beta-lactamase gene and GS-cDNA is glutamine synthetase encoding sequence form vector pRY17).

Figure 12 shows the coverage graph of mapped RNA-seq reads on the Myo LC and HC region of the integrated vector (for RMCE-generated cell lines 1-4). The height of the chart is equal to the coverage at each base without smoothing.

Examples

Example 1:

A) Materials and Methods

1. Vector construction

All vector sequences were synthesized fully sequenced. Puromycin acetyl transferase (PAC), hygromycin phosphotransferase (Hyg) and mAb genes were all gene-optimized and adapted to the codon bias of Cricetulus griseus prior to gene synthesis. The majority of pRY17 (figure 1) was derived from pCB72.3-GA-HC-LC (Kalwy et al. (2006), Mol Biotechnol. 34. pages 151 to 156). Sequences for the wild type FRT (F) and mutant F5 FRT (F5) recombination sequences, present in all vectors, are given in Schlake and Bode (Schlake and Bode, 1994, Biochemistry 33:12746-12751). Template sequences (prior to gene synthesis with or without gene optimization were sourced as follows: In-frame fusion of F recombination sequence and the linker in vector pRY17 (figure 1) were taken from vector pFRT/lacZeo (Invitrogen). The in-frame fusion of the methionine initiation codon and the F recombination sequence in pRY21 (figure 3) was obtained from pcDNAT5/FRT (Invitrogen). In all vectors, the SV40 early promoters used (apart from the one associated with the GS transcription unit, which was derived from pCB72.3-GA-HC-LC), were obtained from the aforementioned Invitrogen vectors. Thymidine kinase (TK) and PAC gene sequences in pRY37 were derived from pWSTK6 (gb:EU530625) and pPUR (Clontech, gb:U07648), respectively.

2. Batch and fed-batch shake flask analysis

For the batch shake flask analysis, cells were seeded at 3 x 10^5 viable cells/mL in 125 mL shake flasks in 30 mL of CD CHO supplemented with various selective agents (as described later) and incubated at 37° C in a humidified 5% CO_2 in air (v/v) orbital shaking incubator at 140 rpm. Conditioned medium was harvested at day 7 of the culture and the antibody concentration in the conditioned medium was determined by Protein A HPLC.

For fed-batch shake flask analysis, cells were seeded at 3 x 10^5 cells/mL in 500 mL shake flasks, each containing 100 mL of proprietary medium and incubated at 37° C in a humidified 5% CO_2 in air (v/v) orbital shaking incubator at 140 rpm. Cells were fed starting on day 3 of the culture with a proprietary feed consisting of mixture of amino acid and trace elements. Daily viabilities and viable cell concentrations were determined using a Vi-CELL™ automated cell viability analyzer. Antibody concentration in the medium was determined by Protein A HPLC starting on day 6 of the culture through to its harvest on day 14.

3. Stability analysis

Cells were sub-cultured alternately every 3 and 4 days in 125 mL shake flask in 30 mL CD CHO supplemented with different selection agents (as described later). At different generation numbers (1 generation is equivalent to 1 population doubling), duplicate fed batch shake flasks were set up as described above. Cell concentration, viability and mAb concentration measurements were collected as described above. If the productivity of a cell line changes by > 30% within 70 generations, then it is considered to be unstable.

4. Flow cytometry for single-cell cloning

Single-cell cloning was performed on a FACS Aria II cell sorter equipped with FACSDiva v6.0 software with an air-cooled laser emitting at 488 nm. Dead cells were excluded in a FSC vs. SSC dot plot, and the doublets were excluded in a FSC width vs. area dot plot. The sorting gate for the live cells was a combination of the two dot plots.

5. Generation of SSI host cells

Transfection of the parental pRY17-expressing cell line with the null vector, pRY37, comprising the second gene coding sequence of interest, namely the selection marker gene, was conducted using FreeStyle™ MAX CHO system (Invitrogen). To this end, 24 h before the re-transfection for RMCE, a selected pRY17-transfected cell line (11A7) was first seeded in FreeStyle™ CHO Expression Medium at 5 x 10^5 cells/mL in a 125 mL shake flask. On the day of transfection, approximately 3 x 10^7 cells at a concentration of 1 x 10^6 cells/mL were co-transfected with 33.75 µg of pOG44 plasmid (Invitrogen, gb:X52327) and 3.75 µg of pRY37 (9:1) with FreeStyle™ MAX reagent in a 125 mL shake flask according to the manufacturer's instructions. Post-transfection, cells were plated into 48-well plates containing proprietary chemical-defined medium supplemented with 25 µM MSX and 7 µg/mL puromycin. Three weeks post-plating, medium from each well containing viable cells were screened for antibody production on a ForteBio using Protein A biosensor. Medium from cell lines with no detectable antibody were advanced to 125 mL shake flasks containing CD CHO medium supplemented with 25 µM MSX and 1 µg/mL puromycin.

6. Generation of cell lines derived from the 10E9 host

RMCE experiments are divided into 3 distinct 'rounds', and are referred to both here and later in the results section below (rounds are defined in figure 6).

In rounds 1 and 2, transfection of the SSI host cell 10E9 with the SSI targeting vector, pRY21 was conducted using FreeStyle™ MAX CHO system (Invitrogen). To this end, 24 h before the re-transfection for RMCE, 10E9 SSI host cells were first seeded at a concentration 5 x 10^5 cells/mL in a 125 mL shake flask containing FreeStyle™ CHO Expression Medium (Invitrogen) supplemented with 25 µM MSX and 1 µg/mL puromycin. On the day of transfection, approximately 3 x 10^7 cells at a concentration of 1 x 10^6 cells/mL were co-transfected with 33.75 µg of pOG44 plasmid (Invitrogen, gb:X52327) and 3.75 µg of pRY21 (9:1) with FreeStyle™ MAX reagent in a 125 mL shake flask according to the manufacturer's instructions. Post-transfection, cells were recovered in a 125 mL shake flask containing 30 mL of FreeStyle™ CHO Expression Medium (Invitrogen) supplemented with 25 µM MSX. 48 h post-RMCE transfectants from round 1 were plated onto 48-well plates containing proprietary chemical-defined medium supplemented with 25 µM MSX, 400 µg/mL hygromycin (positive selection) and 3 µM ganciclovir (negative selection). Four weeks later, the concentration of mAb in medium from wells containing visible foci was determined on a ForteBio Octect using a Protein A biosensor. Cells secreting mAb into medium were expanded and maintained in shake flasks containing CD CHO medium supplemented with 200 µg/mL hygromycin and 25 µM MSX. These cell pools were further evaluated for antibody productivity in batch shake

flask analysis (as described earlier). For stability analysis, MSX was removed from the sub-culture for the condition specified in figure 6.

After recovery for 48 h in shake flasks, round 2 transfectants were seeded at a concentration of 5 x 10^5 cells/mL in a 125 mL shake flask containing CD CHO medium supplemented with 400 µg/mL hygromycin (positive selection). This was followed by the addition of the 3 µM ganciclovir 5 days later as the negative selection. Cells were passaged continuously every 3-4 day in the same medium for 3 weeks in the same shake flask. Surviving cells were single-cell cloned using a FACS Aria II into 96-well plates containing proprietary chemical-defined medium supplemented with 400 µg/mL hygromycin and 3 µM ganciclovir. Three weeks later, the mAb concentration in medium from wells with visible cell growth was determined on a ForteBio Octect using a Protein A biosensor. Clones secreting mAb into the culture medium were expanded and maintained in shake flasks containing CD CHO supplemented with 200 µg/mL hygromycin and 25 µM MSX. These clones were further evaluated for antibody productivity in batch shake flask analysis (as described earlier). For the stability analysis, MSX was removed from the sub-culture for the conditions as specified in the figure 6.

For round 3 transfections, 1 x 10^7 10E9 SSI host cells were co-transfected by electroporation with 45 µg of pOG44 plasmid (Invitrogen, gb:X52327) and 5 µg of pRY21 at 900 µF, 300V. Post-transfection, the cells were seeded into a T-75 flask containing 20 mL proprietary chemical-defined medium. After 48 h, either 200 or 400 µg/mL hygromycin was added to the medium followed by the addition of 3 µM ganciclovir 6 days later. In some cases (as described in the figure 6 legend),

25 µM MSX was added along with the hygromycin selection. For all transfection conditions tested, cells were selected for 3 weeks under these conditions. During this period, conditioned medium in the T-75 flask was exchanged with fresh medium containing the positive and negative drug selection every 3-4 days. Once the viability reached >= 90%, cells were single-cell cloned using a FACS Aria II into 96-well plates containing the same proprietary chemical defined medium described earlier supplemented with either 200 or 400 µg/mL hygromycin with or without 25 µM of MSX (as described in the figure 6 legend). Antibody concentration in conditioned medium, collected from the wells containing growing visible cells, was determined on a ForteBio Octect using a Protein A biosensor. Selected clonal cell lines were expanded and sub-cultured in 125 mL shake flasks containing CD CHO medium supplemented with 200 µg/mL hygromycin and 25 µM MSX. These clones were further evaluated for antibody productivity in batch shake flask analysis (as described earlier). For the stability analysis, both hygromycin and MSX were removed from the sub-culture for the condition specified in figure 6 legend.

7. Data analysis

The time-integral of the area under the growth curve (the time-integral of the viable cell concentration (IVC); 10^6 cells day/mL) was calculated using the method described by Renard et al. (Renard et al. 1988, Biotechnology Letters 10:91-96)

$$IVC = \sum_{k=1}^{n}\left(\frac{Xv_0 + Xv_1}{2}.(t_1 - t_0)\right)$$

where Xv0 = viable cell concentration at first sample (10^6/mL), Xv1 = viable cell concentration at second sample (10^6/mL), t0 = elapsed time at first sample (day), t1 = elapsed time at second sample (day).

8. DNA walking

Seegene DNA Walking SpeedUp™ Kit II was used according to the manufacturer's to provide 3' genome flanking sequence data. Beta-lactamase (bla) gene-specific primers, specific for bla in the 3' arm of the schematic of linearly-integrated pRY17 vector (bla R, figure 3) in GS-CHOK1SV cell line 11A7, TSP1fwd (SEQ ID No. 4), TSP2fwd (SEQ ID No. 5) and TSP3fwd (Seq ID No. 6) were used.

9. Genome sequencing of the 10E9 host

10E9 genomic DNA was fragmented and a paired-end library suitable for HiSeq platform sequencing was prepared using the TrueSeq DNA Sample Preparation kit, following manufacturer's instructions. The library generated was within the expected size range of 300bp to 500bp. QC analysis of the generated library using an Agilent 2100 Bioanalyzer (indicated that the library was of acceptable quality, containing the expected fragment size and yield, for continued sample processing. The library generated was used in the cBot System for cluster generation, following manufacturer's instructions. The flow cells containing amplified clusters were sequenced using 2x100 base pair paired-end sequencing on a Hi-Seq 2000. The reads are mapped to CHO-K1 contigs (Xu X et al. 2011, Nature Biotechnol. 29:735-742) using the Burrows-Wheeler Aligner (BWA) (Li H. and Durbin R. 2009 25:1754-1760).

B) Results

Phase I: Generation of the parental mAb-expressing cell lines

The aim of phase I was to generate a high-producing mAb-expressing GS-CHOK1SV cell line, exhibiting favorable growth characteristics with stable productivity, containing only a single integration locus and the lowest possible number of vector integrants at this locus. A modified Lonza GS 'double gene vector', pRY17 (figure 1), encoding the chimeric mAb, cB72.3 (Whittle et al. 1987, Protein Eng 1:499-505) was designed. In pRY17, the 48 bp wild type F and mutant F5 recombination sequences flank the HC and LC transcription units of the mAb. F5 and F recombination sequences although functionally equivalent, show minimal cross-recombination activity hence permitting efficient and directional RMCE, catalyzed by the FLP recombinase (Schlake and Bode, 1994, Biochemistry 33:12746-12751). Initial GS-CHOK1SV cell lines derived from a pRY17 transfection do not efficiently transcribe or translate the hygromycin resistance gene as there is no initiator methionine codon or promoter immediately upstream of this hybrid gene.

Importantly, the vector-derived GS gene was placed outside of exchangeable cassette so that it would be retained in the genome of resulting cell lines after RMCE. By doing so, any potential perturbation of glutamine metabolism in any derivative cell line was avoided; the parental GS-CHOK1SV cell lines were selected in glutamine-free medium and 50 µM methionine sulfoximine (MSX), in the presence of both endogenous and vector-derived glutamine synthetase expression. A promoterless and translation initiation methionine-deficient (-ATG) hygromycin B phosphotransferase gene were

placed in-between sequence encoding linker (Lnk) and the F recombination sequence (figure 1). The vectors used for RMCE in the second phase of this study contain the SV40 early promoter and the initiation methionine (ATG) required to express hygromycin resistance. A fully-functional hygromycin resistance gene is only created when recombination occurs with the wild type F recombination sequence in the genome (figure 1).

The pRY17 vector containing FRT recombination sequences was introduced into CHOK1SV cells by a conventional cell line development procedure, followed by an intensive screen conducted at key stages of the process to ensure that we isolated cell lines with the best combination of growth and productivity. Additionally, cB72.3 protein derived from the chosen cell lines has to exhibit similar product quality characteristics as a preparation derived from a previous GS-CHOK1SV cell line (Birch and Racher, 2006, Adv Drug Deliv Rev 58:671-685). The HC and LC gene copy numbers from candidate cell lines are preferred to be close to one for each. To this end, three independent electroporations were conducted and each with 50 µg of linearized pRY17 and 1 x 10^7 CHOK1SV cells. Transfectants from all three electroporations were selected in medium containing 50 µM MSX and ~1500 surviving cell pools were screened for antibody production at 3 weeks post-transfection. Eventually a total of 79 clones were evaluated by the 7 day batch shake flask and cB72.3 mAb concentration of all 79 clones was determined. All RMCE derivative cell lines were maintained in medium containing 25 µM MSX, except where stated. From the 79 clones evaluated, 38 were selected for further analysis in fed-batch shake flask culture. The top 6 best-performing clones based on productivity and growth

characteristics were selected (table 1) for genetic characterization (see methods).

Cell Line	Batch	Fed-Batch			
	Product Concentration (mg/L) – day 7	Estimate of IVCC (10^6 cells day/mL)	Product Concentration (mg/L) - day 14	Specific Production Rate (pg/cell.day)	Peak Cell Density (10^6 cells/mL)
1G11	637.73	73.55	3385	47.3	9.00
6B5	273.73	97.89	2029	21.2	13.40
8F10	427.41	125.46	2976	25.0	15.32
11A7	457.88	149.6	3570	25.1	19.13
14D11	378.83	86.32	3251	39.2	10.33
18C11	480.27	90.45	2953	35.0	11.05

Table 1: Growth and antibody production of the top 6 clonal cell lines from Phase 1 in batch and fed-batch shake flask (n=2).

In order to investigate the integration site of the pRY17 in the CKOK1SV genome, metaphase chromosomes from the 6 clones were prepared and probed with DIG-labeled pRY17 (data not shown). Clones 1G11, 6B5, 8F10, 14D11 and 11A7 all appear to have only one integration locus at the telomeric region of an individual chromosome. Clone 18C11 on the other hand seems to have two distinct integration sites and therefore was not selected for further study. To determine the gene copy numbers in each of the cell lines, sonicated genomic DNA was prepared from actively growing cells. For the qPCR analysis, GAPDH was included as the endogenous control and pRY17 was 'spiked' into host cell DNA as the positive control. The gene copy numbers per cell for both the HC and LC were calculated as the ratio of averaged copies to averaged GAPDH copies. As shown in table 2, out of the 5 clones analyzed, 11A7 has the lowest HC and LC gene copy numbers. Southern blot analysis of the genomic DNA revealed that both HC and LC can be detected in

all 5 clones and the intensity of bands reflects the qPCR-determined gene copy numbers (data not shown).

Cell Line	HC average copies/cell	LC average copies/cell
1G11	15.58	13.95
6B5	44.46	47.91
8F10	40.46	38.41
11A7	6.57	3.57
14D1	9.75	9.31

Table 2: Gene copy analysis of the cB72.3 HC and LC in 5 of the 6 clonal cell lines using qPCR

Out of the 5 clones selected that entered the present stability study (table 3), 11A7 maintained similar productivity over 7 months (220 generations), whereas other clones showed gradual productivity loss during the first 3 months of the study (80 generations). Taken together, 11A7 not only has one of the best combinations of good growth and productivity profiles, but also has the lowest gene copy number with a single integration site. Importantly, 11A7 is the most stable clone out of the 6 in terms of productivity. Most importantly product quality was comparable after 220 generations. 11A7 was chosen as the parental clone for the first round of RMCE: Phase II.

Cell line	Antibody production (mg/L) at given generation number									
	10	40	80	100	120	140	160	180	200	220
1G11	2809.00	2977.00	2675.86	2553.18	2464.09	2047.00	-	-	-	-
11A7	2946.00	2776.82	2546.47	2464.99	2395.90	2421.00	2425.00	2498.00	2496.00	2676.00
6B5	1548.00	1791.94	1580.33	-	-	-	-	-	-	-
8F10	2433.00	2593.00	2760.20	-	-	-	-	-	-	-
14D11	2629.00	2521.08	2358.44	-	-	-	-	-	-	-

Table 3: Stability analyses of the top 5 SSI clonal cell lines. The top 5 SSI clonal cell lines were continuously cultured in shake flasks for various numbers of generations in the presence of MSX. At different generation numbers as indicated, all 5 SSI clonal cell lines were analysed by fed-batch shake flask for antibody production. (n =2).

Phase II - Generation of SSI host cells

Although it is entirely possible to design a targeting vector that could swap the original mAb transcription units in 11A7 for those of a new mAb, it is preferred that the original was completely excised from the genome. To do this an additional null targeting vector, pRY37 (figure 2) was designed that does not encode mAb genes. Instead it encodes positive and negative selection markers, puromycin acetyl transferase (PAC) and thymidine kinase (TK), respectively, and these were flanked with F and F5 recombination sequences in the same orientation as pRY17 (figure 1). After RMCE, these markers are expected to replace the original cB72.3 mAb in the same locus in the original 11A7 genome. PAC is required to select for cell lines that have undergone RMCE with pRY37. TK converts the prodrug ganciclovir into a toxic, phosphorylated nucleotide analogue (Wood and Crumpacker, 1996, New England Journal of Medicine 335, pages 721 to 729). This is important for the application of negative selection later in phase III: Ganciclovir selection enriches the RMCE pool for cells that have undergone legitimate RMCE, by killing those that have not.

Of the surviving cell line pools that were negative for mAb expression, one cell line, 136-A4 was chosen for further characterization by Southern blot analysis (data not shown). It confirmed the presence of TK in the 136-A4 genome. Restriction mapping indicated the presence of only two copies of pRY37 in the “hot-spot” and was confirmed by subsequent genome sequencing of the daughter clone 10E9. The copy number is substantially lower than pRY17 found in 11A7 (table 2). To obtain a homogenous SSI host, we single-cell

cloned 136-A4 using FACS Aria II and obtained growth profiles of 26 clonal derivatives. Out of these, two clonal derivatives with the best growth profiles, 10E9 and 8C8, were selected for further characterization by northern blot analysis. Northern blot analysis of RNA from these daughter clones confirmed the absence of cB72.3 HC and LC mRNAs (data not shown). Taken together, these results show that 10E9 is a suitable candidate host cell line for RMCE for testing in phase III.

Phase III – RMCE with Myo mAb targeting vector

In order to demonstrate the utility of the new SSI host cell line 10E9, a targeting vector, pRY21 was designed (figure 3). This vector contains HC and LC genes for the Myo mAb, flanked by F and F5 recombination sequences in the same orientation as found in both pRY17 (phase I) and pRY37 (phase II) vectors. It also contains a SV40 early promoter (SV40E) upstream of a methionine initiation codon fused in-frame to the F recombination sequence (ATG + F). When legitimate RMCE occurs, the ATG + F sequence is placed in frame with the promoterless and translation initiation methionine-deficient (-ATG) hygromycin B phosphotransferase gene. Three rounds of co-transfection of 10E9 cells with pRY21 and pOG44 were performed. In the first two rounds cells were transfected with Free style™ MAX CHO system, whereas in round 3 cells were transfected by electroporation. Additionally, round 1 was co-selected with ganciclovir and hygromycin as pools; round 2 was sequentially selected first with 400 µg/mL of hygromycin followed by ganciclovir as pools and were then subsequently single-cell cloned using a FACS Aria II; round 3 was sequentially selected first with either 200 or 400 µg/mL of hygromycin followed by ganciclovir as pools and then were subse-

quently single-cell cloned using a FACS Aria II (figure 6). In round 3, MSX was absent from the culture medium in two of the test conditions.

The productivity data obtained from different pools in round 1 is similar, suggesting that cell line members of each pool are likely to have similar productivities. The range of productivities from the pools is much narrower than that of either clonal or non-clonal cell lines from a random integration process (figure 6). When comparing the clonal cell lines isolated in different selection conditions from rounds 2 and 3, it appears that the highest-producing clonal cell lines reside in the group selected in the absence of MSX as compared to the ones with MSX. In the cell lines grown in the absence of MSX, there appeared to be a difference in the productivities obtained with different concentrations of hygromycin in the selective medium which was not apparent when MSX was present in the medium

Initially, in phase I a very stable GS-CHOK1V cell line, 11A7, was isolated that is stable up to 220 generations. This stability trait could be inherited in derivative cell lines generated by RMCE in phase III. Accordingly, 3 cell pools from round I from phase III were evaluated in an extended stability study under two different conditions (Table 4).

Generation Number	Antibody concentration (mg/L) in fed-batch cultures of cell lines					
	70C2		72A3		74B5	
	(+) / (+)	(-) / (+)	(+) / (+)	(-) / (+)	(+) / (+)	(-) / (+)
10	1240	1140	1510	1430	1310	1170
40	1070	1140	1400	1250	1370	1190

70	1100	1170	1400	1200	1370	1240
100	1030	1270	1700	1490	1700	1400

Table 4: Round 1: Three selected cell lines were expanded into shake flasks and cultured continuously at two different conditions; + hygromycin/+ MSX (+/+) or – MSX/+ hygromycin (-/+). At various time points, duplicate fed-batch shake flask cultures were setup from the continuous cultures of all 3 SSI cell lines in both conditions and the concentration of Myo mAb in medium was determined after 14 days.

Regardless of the conditions, all three pools tested met the criteria for a stable cell line. Further, a total of 12 clonal cell lines, 6 each from round 2 and 3, in the same type of stability study (tables 5 and 6, respectively). It was found that all 12 clonal cell lines retained the stability trait under selection. Interesting was that the 6 clonal cell lines from round 3 (table 4) were stable even without the presence of any selective agents. This has profound implications for manufacture of biopharmaceuticals.

Generation Number	Antibody concentration (mg/L) in fed-batch cultures of cell clones											
	11F6		11F11		12C3		13F7		13G8		15E9	
	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)
10	1470	1390	1920	1860	1480	1360	1960	1870	2020	2020	2150	2100
40	1300	1420	1870	1680	1770	1590	1850	1860	1940	2010	2020	2100
70	1350	1350	1870	1650	1770	1600	1850	1760	1840	1890	2060	2060
100	1200	1580	2180	1930	2160	1790	2094	1950	1970	2020	2500	2360

Table 5: Round 2: Six single-cell clones were continuously cultured at two different conditions; + hygromycin/+ MSX (+/+) or – MSX/- hygromycin (-/-) and were periodically analyzed by fed-batch culture as described for Round 1 (table 4).

Generation Number	Antibody concentration (mg/L) in fed-batch cultures of cell clones											
	2H4		4H7		A5B8		A7D5		A8A10		A8G1	
	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)	(+) / (+)	(-) / (-)
10	2400.0	2400.0	2200.0	2400.0	1900.0	2100.0	2200.0	2300.0	1300.0	1400.0	1400.0	1500.0
40	2460.0	2440.0	2330.0	2240.0	2100.0	2070.0	2320.0	2320.0	1580.0	1570.0	1490.0	1500.0
70	2080.0	2150.0	1980.0	2070.0	1910.0	1930.0	1990.0	2120.0	1490.0	1610.0	1360.0	1380.0
100	2389.0	2359.0	2655.0	2510.0	2215.5	2276.5	2343.0	2536.0	1760.0	2009.0	1637.5	1656.5
130	2477.5	1937.5	3401.0	2505.0	2392.0	1970.0	2496.0	2311.0	1760.0	1870.0	1552.0	1029.0

Table 6: Round 3: Six single-cell clones were continuously cultured at two different conditions; -MSX/+hygromycin (-/+) or – MSX/- hygromycin (-/-) and were periodically analysed by fed-batch culture as described for Round 1 (table 4).

Characterization of genomic sequences flanking the "hot-spot"

3' flanking sequence

500 bp 3' flanking sequence (SEQ ID No. 7) sequence derived from Seegene DNA walking from bla R (figure 3) was used to blast-search the contigs of a CHO-K1 genome sequencing project (Xu X et al. 2011, Nature Biotechnol. 29:735-742), publically available in the NCBI databank. A unique region located on unplaced genomic scaffold, scaffold1492 (accession number, JH000254.1, identical to NW_003613833.1) was found on the minus strand from 1760466-1760965. The 500 bp sequence (SEQ ID No. 7) was extendend to 2000 bp (Seq ID No. 8) based high coverage mapping of the 10E9 reads to the scaffold1492 (data not shown).

The 3' flanking sequence, identified by Seegene DNA walking (see methods) was used to blast search the contigs of a CHO-K1 genome sequencing project (Xu X et al. 2011, Nature Biotechnol. 29:735-742), publically available in the NCBI databank. Using this data and Illumina HiSeq genome sequence data obtained from the 10E9 SSI host cell line, a unique region located on unplaced genomic scaffold, scaffold1492 (accession number, JH000254.1, identical to NW_003613833.1) was found. This was found to be located within a predicted Fer1L4 (fer-1-like 4) gene (NCBI Gene ID: 100755848) in scaffold1492 on the minus strand (scaffold 1492 nucleotide number 1,746,191 to 1,781,992; 35,802 nucleotides in total). The 5' flanking sequence appears to be located between exons 39 and 40 whilst the 3' flanking sequence appears to be located between exons 28 and 29 (see figure 7).

5' flanking sequence

Illumina reads from 10E9 genomic DNA were mapped to pRY17 (SEQ ID No. 1) using the Burrows-Wheeler Aligner (BWA). Through inspection of the mapping, it was found that multiple unpaired reads (black arrows in the Figure 8) mapped to the 5' end of a 3' portion of the cB72.3 HC. This analsyis suggests at least one fragment of cB72.3 HC (with 214 bp deleted from the 5' end) remains in the 10E9 "hot-spot". Northern blot analsyis (data not shown) showed that there was no cB72.3 LC or HC mRNA in 10E9 cells, suggesting that the truncated HC in cB72.3 is non-functional. The equivalent paired reads were then used to blast-search the the contigs of a CHO-K1 genome sequencing project (Xu X et al. 2011, Nature Biotechnol. 29:735-742). All of them aligned to the same location (1750048-1750183) on unplaced genomic scaffold, scaffold1492 (accession number, JH000254.1, identical to NW_003613833.1). This allowed the 5' flanking sequence to be extended to a maximum of 822bp (SEQ ID No. 9).

Example 2:

A) Materials and Methods

1. Southern blot

5-10 µg of genomic DNA, isolated from passages 2 and 4 of each clone and purified using Blood & Cell Culture DNA Maxi Kit from QIAGEN (Qiagen), was digested with restriction endonuclease(s) for 15 h at 37 °C. The digested DNA was extracted twice with an equal

volume of a phenol:chloroform:isoamyl alcohol mixture, pH8.0 (1:1 v/v) followed by chloroform alone and ethanol-precipitated prior to electrophoresis on 0.7% (w/v) agarose gel run in either 0.5 x TBE (50 x TBE: Lonza) or 1 x TAE (40 mM Tris, pH 7.7, 2.5 mM EDTA) buffer. The gel was transferred onto Hybond-N membrane (Amersham) using a vacuum manifold essentially according to manufacturer's instructions (Appligene, Pharmacia). The Hybond-N membranes were UV-fixed, pre-hybridized in either hybridization buffer containing 5 x Denhardt's prepared from 50 x stock solution (Sigma), 6 x SSC (1 x SSC: 0.15 M sodium chloride, 15 mM sodium citrate), and 10 % (w/v) SDS or Rapid-hyb Buffer (GE healthcare) alone.

TK probes were generated in PCRs using the following primer sets:

TK-forward: 5'-AGATCACCATGGGCATGCCTTAC-3' (SEQ ID No. 10);

TK-reverse: 5' AACACGTTGTACAGGTCGCCGTT-3' (SEQ ID No. 11);

The vector pRY37 was used as a template for the probe-generating PCR and the cycling conditions were: 15 ng template/50µl reaction; Taq DNA Polymerase (Roche); 94 °C 2 min, 30 cycles of 94°C for 30s, 55°C for 1 min and 72°C for 30s, final extension at 72°C for 7 min. 25 ng of PCR product was labeled with [γ-32P] dCTP (111 TBq/mmol, Perkin Elmer) using the Megaprime Kit and purified on a nick-translation column (Amersham). Hybridizations were performed in the same pre-hybridization buffer for 2 – 20 h at 65 ºC. Post-hybridization, membranes were washed to a final stringency of 0.1 x SSC, 0.1 % (w/v) SDS at 65 °C. Blots were exposed to a storage

phosphor screen (Bio-Rad); exposed screens were imaged using a Personal Molecular Imager (PMI) System (Bio-Rad).

2. Mapping and alignment

The paired end reads in FASTQ format are the input for mapping to genomics templates. Vector sequences and CHOK1SV assembly are indexed as the templates to be mapped to. The paired end reads are aligned to the templates using Bowtie2 (Langmead B, & Salzberg SL, 2012, Nature methods, 9 (4), 357-9) with the default parameters (-D 5 -R 1 -N 0 -L 25 -i S,1,2.00) for very fast local alignment. Coverage is normalized as the <raw coverage>*500M / <number of reads> in order to compare across different samples.

3. Identification of Integration Sites

2 X 100 paired end reads of 10E9 SSI host strain were sequenced using Illumina Hi-Seq 2000 at an average coverage of 40 X. The sequence reads were mapped to vector pRY17 which is the first vector integrated into the CHOK1SV genome. Reads covering integration sites are termed chimerical reads because they contain sequence that maps to both the CHOK1SV genome and also to integrated vector sequence. Because the mapping is performed by local alignment, the chimerical reads have characteristics of partial match to vector sequences with overhang tails which could map to genomic sequence. In addition to the chimerical reads, there other reads where one end of a paired read maps to vector sequence fully and the

other end maps to genomics sequence. These read pairs are called discordant read pairs. The overhang tail sequences and unmapped reads from discordant read pairs are collected and used to search against CHOK1SV genome assembly using blast to identify the flanking sequence of integration sites based on sequence similarity.

4. Landing pad

The structure of landing pad (exogenous sequences introduced into the hot-spot that contain recombination sites for the integration of expression cassettes of genes of interest by RMCE) was based on both the Southern blot analysis and whole-genome re-sequencing (WGRS) analysis data of the 10E9 cell line (figure 9). 10E9-derived reads mapped to the vector sequence using the same algorithm as used to map the integration sites. Chimerical reads are also observed at duplication, deletion or insertion sites. Corrections to the putative landing pad sequence were made based on the close investigation of the sites where chimerical reads arise.

5. Quantification of RNA-seq analysis

The template used to map the reads was constructed using a 'one-copy' model of the landing pad derived from whole genome re-sequencing (see above). RNA-seq sequencing reads were mapped to the template by BWA using default parameters. The read counts on LC and HC were normalized to the RPKM measure, reads per

kilobase transcriptome per million mapped reads, by the following formula:

$$RPKM = \frac{number\,of\,reads\,overlapping\,with\,the\,exons}{\frac{total\,number\,of\,mapped\,reads}{10^6} \times \frac{total\,length\,of\,all\,exons}{10^3}}$$

The number of reads overlapping with the exons is obtained for each interval using bedtools (Quinlan AR and Hall IM, 2010, Bioinformatics. 26, 6, pp. 841–842) as the following command, bedtools coverage -abam <bam file> -b <intervals in bed>

B) Results

Structure of the landing-pad in 10E9 SSI host cell

A model of the structure of the landing pad within the Fer1L4 hotspot was inferred from the expected RMCE events occurring during the creation of cell line 10E9 from 11A7 using the null targeting vector pRY37 (figure 4). This model is termed the 'single-copy' model and was consistent with the WGRS data from cell line 10E9 (figure 9A). However, the Southern blot data suggested that the model should actually contain two copies (figure 9B). This 'two-copy' model was refined on the basis of the data and was used to help interpret mAb-producing cell lines derived by RMCE in the 10E9 host. The 'two-copy' model allows us to explain why either one or two copies of antibody transcription units can be incorporated into the landing pad by RMCE with the targeting vector pRY21.

Flanking sequences in the RMCE derivative cell lines

Four 10E9-derived recombinant cell lines were created expressing an anti-Myostatin monoclonal antibody (Myo) by RMCE (using targeting vector pRY21) and the 5' and 3' flanking sequences in each were determined, using the methods previously described. During the process of RMCE, a consistent genomic rearrangement occurs generating a new 3' flanking sequence in the derivative cell lines (figure 10). The 5' flanking sequence in the RMCE derivative cell lines are the same as 10E9, with the integration site at nucleotide 1750049 (on the unplaced CHO-K1 Scaffold1492 scaffold (accession number JH000254.1, identical to NW_003613833.1). However, the 3' flanking sequences are now different: Whereas the 3' integration site in the 10E9 SSI host cell line is at nucleotide 1760965, in all the RMCE derivative cell lines it is now found at nucleotide 1435427 (in the aforementioned scaffold, figure 10).

Estimation of the copy number of integrated cassette in RMCE-generated cell lines

The sequencing reads from each of their genomes were mapped to a model where one copy was integrated into the hot-spot. The number of Myo copies was derived from the mean average coverage on LC and HC region. The mean coverage for these four cell lines are 41, 34, 18, 27 on LC and 32, 27, 14, 19 on HC region, respectively. The coverage data indicate that for high producers, there may be at least one more copy of HC and LC (table 7). The coverage data is shown graphically in figure 11.

Cell Line	qP pg/(cell.day)	Coverage	
		LC	HC
1	20	41 (2)	32 (2)
2	19	34 (2)	27 (2)
3	10.2	18 (1)	14 (1)
4	10	27 (1)	19 (1)

Table 7: Sequence read coverage and specific production rate (qP) of Myo producing RMCE cell lines. Copy number is indicated in brackets next to the coverage value.

Based on this observation, a new model of the post-RMCE locus was generated in which an extra copy of Myo was included. This was achieved by inserting another copy of the fragment spanning from the beginning of the first wFRT site to the beginning of the second wFRT site (both indicated by asterisks in figure 11) just before the start of the second wFRT. Using this 'two-copy' model we could fully account for the number of observed reads in cell lines with high qP when they were re-mapped to it.

Further evidence for a 'two-copy' model in high qP Myo-producing cell lines was obtained from RNA-Seq data from each cell line. The sequencing reads from RNA-Seq from RNA derived from each Myo cell line were mapped to one the original 'one-copy' model (figure 12) using by BWA (Li H. and Durbin R. 2009, Bioinformatics, 25:1754-60) run with default parameters. The number of Myo copies was derived by calculating RPKM values (Mortazavi et al.(2008), Nature Methods 5, 621-628) for the LC and HC regions. The RPKM values are 43135, 40059, 23204, 29334 in LC and 38572, 32384, 17878, 20751 in HC region, respectively. These data also confirm the ‘two-copy’ model proposed for high producers we derived from whole genome sequencing.

Claims

1. A site-specific integration (SSI) host cell comprising an endogenous Fer1L4 gene, wherein an exogenous nucleotide sequence is integrated in said Fer1L4 gene.

2. The SSI host cell of claim 1, wherein the exogenous nucleotide sequence comprises at least two recombination target sites.

3. The SSI host cell of claim 2, wherein the at least two recombination target sites flank a gene coding sequence of interest.

4. The SSI host cell of claim 1, wherein the exogenous nucleotide sequence comprises a gene coding sequence of interest.

5. The SSI host cell according to claim 3 or 4, wherein the gene coding sequence of interest comprises one or more of a gene encoding a selection marker, a detectable protein, an antibody, a peptide antigen, an enzyme, a hormone, a growth factor, a receptor, a fusion protein or other biologically active protein.

6. The SSI host cell according to claim 2, wherein the recombination target site is a FRT site or a lox site.

7. The SSI host cell according to claim 5, wherein the selection marker is a GS selection marker, a hygromycin selection marker, a puromycin selection marker or a thymidine kinase selection marker.

8. The SSI host cell according to any one of the preceding claims, wherein the host cell is a mouse cell, a human cell or a CHO host cell, a CHOK1 host cell or a CHOK1SV host cell.

9. The SSI host cell according to any one of the preceding claims, wherein the nucleotide sequence of the Fer1L4 gene flanking the integrated exogenous nucleotide sequence is selected from the group consisting of SEQ ID No. 7, 8, 9 and homologous sequences thereof.

10. The SSI host cell according to any one of the preceding claims, wherein the exogenous nucleotide sequence replaces a portion of the Fer1L4 gene.

11. The SSI host cell according to claim 6, wherein the host cell comprises a FRT site which is a wild type FRT site or a mutant FRT site.

12. The SSI host cell according to claim 2, wherein the recombination target sites comprise at least one wild type FRT site and at least one mutant FRT site.

13. A method for producing an SSI host cell comprising the steps of

a) providing a cell comprising an endogenous Fer1L4 gene, wherein an exogenous nucleotide sequence is integrated in said Fer1L4 gene, and wherein the exogenous nucleotide sequence comprises at least two recombination target sites, flanking at least one first gene coding sequence of interest; subsequently

b) transfecting the cells provided in step a) with a vector comprising a first exchangeable cassette, the cassette comprising at least two matching recombinant target

sites, flanking at least one second gene coding sequence of interest, namely a selection marker gene, subsequently

c) effecting a site-directed recombination-mediated cassette exchange and subsequently

d) selecting transfected cells expressing second gene coding sequence of interest, preferably the selection marker gene, so as to obtain the SSI host cell comprising the first exchangeable cassette stably integrated in its genome.

14. The method of claim 13, further comprising i) transfecting the SSI host cell with a vector comprising at least one second exchangeable cassette, which cassette comprises at least two matching recombination target sites flanking at least one third gene coding sequence of interest and at least one selection marker, ii) effecting a site-directed recombination-mediated cassette exchange so as to obtain a SSI host cell comprising the third gene coding sequence of interest, iii) allowing the SSI host cell obtained in step ii) to express the third gene coding sequence of interest and iv) recovering the product of the third gene coding sequence of interest.

15. An isolated polynucleotide molecule comprising at least one nucleotide sequence selected from the group consisting of nucleotide sequences as given in SEQ ID No. 7, 8, 9 and homologous sequences thereof.

16. A vector containing a polynucleotide of claim 15.

17. A host cell comprising the nucleotide molecule according to claim 15 or the vector of claim 16.

18. The host cell of claim 17, wherein the host cell comprises a gene coding sequence of interest within or adjacent to the polynucleotide molecule of claim 15 or the vector of claim 16.

19. A method for producing a cell or cell line, in particular a high producer cell or cell line with a high productivity stability, wherein a Fer1L4 nucleotide sequence is integrated in a vector comprising at least one gene coding sequence of interest, wherein subsequently the vector is stably transfected into a cell or cell line and wherein subsequently stably transfected cells or cell lines are selected and obtained.

20. The method of claim 19, wherein the Fer1L4 nucleotide sequence is a nucleotide sequence selected from the group consisting of nucleotide sequences as given in SEQ ID No. 7, 8, 9 and homologous sequences thereof.

21. A method for the production of a product of a gene coding sequence of interest comprising cultivating a host cell or cell line produced according to the method of claims 13 or 19 in a suitable medium and recovering the product therefrom.

22. A method of producing an SSI host cell, the method comprising introducing an exogenous nucleotide sequence into an endogenous Fer1L4 gene in a cell.

23. The method of claim 22, wherein the exogenous nucleotide sequence is introduced by homologous recombination between the

Fer1L4 gene and a polynucleotide, wherein the polynucleotide comprises

a) a first nucleotide sequence homologous to a first portion of the Fer1L4 gene,

b) the exogenous nucleotide sequence, and

c) a second nucleotide sequence homologous to a second portion of the Fer1L4 gene.

24. The method of claim 22, wherein introduction of the exogenous nucleotide sequence is facilitated using a viral vector (e.g., an adeno-associated virus vector which mediates homologous recombination) or an exogenous nuclease (e.g., a zinc finger nuclease, a transcription activator-like effector nuclease, or an engineered meganuclease).