EFFICIENT SELECTIVITY OF RECOMBINANT PROTEINS

BACKGROUND

Sequence listing

[0001] This application incorporates by reference the Sequence Listing submitted in Computer Readable Form as file 8700WO_ST25.txt created on August 3, 2015 (75,769 bytes).

Field of the Invention

[0002] The invention provides for expression of recombinant proteins in mammalian cells in a consistent and efficient manner. In particular, the invention includes methods and compositions for improved expression of proteins in mammalian cells by employing mammalian selection markers. The invention includes methods that facilitate selectivity and enhanced expression copies as well as protein yield of recombinant proteins in mammalian cells, and methods of using such expression systems.

Description of Related Art

[0003] The development of cellular expression systems is an important goal for providing a reliable and efficient source of a given protein for research and therapeutic use. Recombinant protein expression in mammalian cells is often preferred for manufacturing therapeutic proteins due to, for example, the ability of mammalian expression systems to appropriately post-translationally modify recombinant proteins.

[0004] Various vectors are available for expression in mammalian hosts, each containing selection markers that enable ease of isolation of recombinant protein-expressing cells during cell culture. Selectable marker genes (SMGs) are utilized in such systems because they confer a selective advantage for cells expressing the protein of interest, however SMGs must be optimized for their phenotypic neutrality, efficiency and versatility, among other reasons.

[0005] Despite the availability of numerous vectors and expression systems hosting SMGs, the expression of a recombinant protein achieved in mammalian systems is often unsatisfactory, whether in quantity or quality or both. The biological “fingerprint” of a molecule, for example post-translational modifications like glycosylation, is of particular importance in defining the molecule’s utility and efficacy in the development of a recombinant protein therapeutic (Cumming, D.A., 1990, *Glycobiology*, 1(2):115-130). SMGs that do not negatively impact the biological properties of an expressed protein of interest are particularly advantageous.

[0006] Most SMGs are of bacterial origin and impart other disadvantages for use in mammalian systems due to growing concern for the risk of horizontal transfer of bacterial antibiotic resistance genes to environmental bacteria (Breyer, D. et al., 2014, *Critical Reviews in Plant Sciences* 33:286-330). Elimination of use of bacterial antibiotic resistance genes could have positive effects on consumer acceptance and alleviating such perceived risks.

[0007] Gene-engineered autologous cells are rapidly becoming a clinical success (see *e.g.*

Kershaw, M.H. et al., 2013, *Nature Reviews: Cancer* 13:525-541). The choice and design of vectors for genetic modifications in human autologous cell products is critical, especially since the unwanted introduction of non-human components to a human autologous cell could have serious consequences for patient safety (Eaker, et al. 2013, *Stem cells Trans. Med*. 2:871–883; first published online in SCTM*EXPRESS* October 7, 2013). A vector system having only components of mammalian origin, rather than bacterial, would be advantageous for use in patient-specific T cells for adoptive immunotherapy.

[0008] Thus it is desirable to introduce mammalian selectivity genes, especially those that give the transformed cells a phenotypic or metabolic advantage in expression systems for the production of mammalian proteins of interest. Moreover, a cell line that reliably expresses sufficiently high levels of a therapeutic protein, and appropriately and consistently modifies the therapeutic protein post-translationally, is highly desirable. Accordingly, there is a need in the art for improved mammalian expression systems.

BRIEF SUMMARY

[0009] The use of a mammalian tunicamycin (Tn) resistance gene as a selectable marker in a mammalian expression system can increase efficiency and copy number of transfectants. It has been observed that the use of a Tn resistance gene operably linked to a gene of interest creates selective pressure on a population of mammalian cells thereby increasing random integration of the transfectant (*i.e*. gene of interest). It is understood that selectable marker systems may foster selection of desired transfectants, however the methods of the invention impart an unexpected increase in both efficiency and random integration of the gene of interest, as well as reliable biological qualities of the desired protein. The compositions and methods of the invention thus allow the advantageous selection of qualitatively favorable post-translational modifications for expressed proteins.

[0010] In one aspect, the invention provides an isolated cell comprising a mammalian tunicamycin (Tn)-resistance gene encoding a protein having at least 93% identity to the amino acid sequence of SEQ ID NO:3, operably linked to a gene of interest (GOI) and at least one regulatory element.

[0011] In another aspect, the invention provides a method of producing a recombinant protein of interest (POI), wherein the method comprises: providing a mammalian host cell encoding a nucleic acid molecule comprising (i) a mammalian tunicamycin (Tn)-resistance gene and (ii) a gene encoding the POI; culturing the cell in the presence of a first concentration of Tn; isolating a cell population expressing at least one copy of the Tn-resistance gene; culturing the cell population in the presence of increasing concentrations of Tn, wherein increasing the concentration of Tn increases production of the POI; and isolating the POI from the cell culture.

[0012] In yet another aspect, the invention provides a method of glycosylating a N-glycan protein substrate, wherein the method comprises: providing a mammalian host cell encoding a nucleic acid molecule comprising a mammalian tunicamycin (Tn)-resistance gene operably

linked to a gene encoding the protein substrate in need of glycosylation; culturing the cell in the presence of a first concentration of Tn; isolating a cell population expressing at least one copy of the Tn-resistance gene; culturing the cell population in the presence of increasing concentrations of Tn, wherein increasing the concentration of Tn increases production of the POI; and isolating the protein substrate from the cell culture.

[0013] In some embodiments of the methods, the Tn-resistance gene is operably linked to the gene encoding the POI, and the gene encoding the POI is operably linked to at least one regulatory element.

[0014] In some embodiments, the Tn-resistance gene is exogenously added to the cell. In other embodiments, the Tn-resistance gene encodes the protein having at least 93% identity to the amino acid sequence of SEQ ID NO:3. In other embodiments, the Tn-resistance gene encodes the protein having at least 94% identity to the amino acid sequence of SEQ ID NO:3. In some embodiments, the Tn-resistance gene encodes the protein having at least 93% identity to the amino acid sequence of SEQ ID NO:4. In still other embodiments, the Tn-resistance gene encodes the protein having at least 94% identity to the amino acid sequence of SEQ ID NO:4.

[0015] In some embodiments, the mammalian Tn-resistance gene comprises a Chinese hamster (*Cricetulus griseus*) Tn-resistance gene. In other embodiments, the mammalian Tn-resistance gene comprises a human Tn-resistance gene.

[0016] The Tn-resistance gene may also comprise the nucleic acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16 and SEQ ID NO:17.

[0017] In certain embodiments of the aforementioned inventions, the mammalian Tn-resistance gene comprises a nucleic acid sequence having at least 92% identity to the nucleic acid sequence of SEQ ID NO:2. In some embodiments, the mammalian Tn-resistance gene comprises a nucleic acid sequence having at least 92% identity to the nucleic acid sequence of SEQ ID NO:12.

[0018] At least one regulatory element operably linked to the Tn-resistance gene is provided in the isolated cell of the invention, wherein the regulatory element includes, but is not limited to a promoter, ribosome-binding site, and enhancer. In still another embodiment, the GOI is operably linked to a promoter. In another embodiment, the GOI is operably linked to a ribosome-binding site, such as an IRES.

[0019] In some embodiments, the isolated cells and methods of the invention further comprise a second gene of interest (GOI), whereas the GOI encodes the protein of interest (POI). In one embodiment, the gene of interest (GOI) is an exogenously added GOI. In another embodiment, the exogenously-added GOI is a human gene. In yet another embodiment, the regulatory element is an exogenously added regulatory element.

[0020] In other embodiments, the first and/or second GOI encodes a POI including, but not limited to an antibody heavy chain, antibody light chain, antigen-binding fragment, and/or Fc-fusion protein.

[0021] In another embodiment, the first GOI and the second GOI are independently selected from the group consisting of a gene encoding for an antibody light chain or antigen-specific fragment thereof, an antibody heavy chain or antigen-specific fragment thereof, an Fc-fusion protein or a fragment thereof, and a receptor or ligand-specific fragment thereof. In one embodiment, a recombinase recognition site is present between the first GOI and the second GOI. In other embodiments, the invention further provides a recombinase recognition site 5' to the first GOI and a recombinase recognition site 3' with respect to the second GOI.

[0022] In still another embodiment, the GOI encodes a glycoprotein selected from an antibody light chain or antigen-binding fragment thereof, an antibody heavy chain or antigen-binding fragment thereof, an Fc-fusion protein or a fragment thereof, a ligand, and a receptor or ligand-binding fragment thereof.

[0023] The isolated, non-naturally occurring cells of the invention may be derived from a eukaryotic cell. In one embodiment, the cell is a mammalian cell. In some embodiments, the isolated cell is an *ex vivo* human cell. In other embodiments, the cell is selected from the group consisting of CHO (e.g. CHO K1, DXB-11 CHO, Veggie-CHO), COS (e.g. COS-7), lymphocyte, stem cell, retinal cell, Vero, CV1, kidney (e.g. HEK293, 293 EBNA, MSR 293, MDCK, HaK, BHK21), HeLa, HepG2, WI38, MRC 5, Colo25, HB 8065, HL-60, Jurkat, Daudi, A431 (epidermal), CV-1, U937, 3T3, L cell, C127 cell, SP2/0, NS-0, MMT cell, tumor cell, and a cell line derived from an aforementioned cell. In certain embodiments, the isolated cell of the invention is a CHO-K1 cell, a lymphocyte, retinal cell, or stem cell.

[0024] In one embodiment, the first concentration of Tn is 1 µg/mL. In another embodiment, the increasing concentrations of Tn comprises a second and third concentration of Tn.

[0025] In some embodiments, the second concentration is greater than the first concentration of Tn, and the third concentration is greater than the second concentration of Tn. In certain embodiments, the second concentration of Tn is 2.5 µg/ml, and the third concentration is 5 µg/mL.

[0026] In still other embodiments, the increasing concentrations of Tn comprises a second concentration of Tn, wherein the second concentration of Tn is 2.5 µg/ml or 5 µg/mL.

[0027] Any of the aspects and embodiments of the invention can be used in conjunction with any other aspect or embodiment of the invention, unless otherwise specified or apparent from the context.

[0028] Other objects and advantages will become apparent from a review of the ensuing detailed description.

BRIEF DESCRIPTION OF THE FIGURES

[0029] Figure 1 illustrates a schematic diagram of the operative expression cassette in a cloning vector construct, used for introduction of the nucleic acid sequence encoding a gene of interest, for example eGFP, into a cell genome. SV40 Promoter: Simian virus 40 Promoter; GPT: GlcNAc-1-P transferase (*e.g.* CHO-GPT, SEQ ID NO:2; or hGPT, SEQ ID NO:12); IRES:

internal ribosomal entry site; eGFP: enhanced Green Fluorescent Protein; SV40polyA: Simian virus 40 polyA.

[0030] Figures 2A to 2C represent an alignment of mammalian GPT amino acid sequences, namely human (GPT_HUMAN; UniProtKB Accn. No. Q9H3H5; SEQ ID NO:4), Rhesus macaque (GPT_MACMU; UniProtKB Accn. No. F6TXM3; SEQ ID NO:5), chimpanzee (GPT_PANTR; UniProtKB Accn. No. H2R346; SEQ ID NO:6), dog (GPT_CANFA; UniProtKB Accn. No. E2RQ47; SEQ ID NO:7), guinea pig (GPT_CAVPO; UniProtKB Accn. No. E2RQ47; SEQ ID NO:8), rat (GPT_RAT; UniProtKB Accn. No. Q6P4Z8; SEQ ID NO:9), and mouse (GPT_MOUSE; UniProtKB Accn. No. P42867; SEQ ID NO:10) compared to Chinese hamster (GPT_CRIGR; UniProtKB Accn. No. P24140; SEQ ID NO:3) GPT amino acid sequences.

[0031] Figures 3A and 3B exemplifies how protein optimization can be achieved using the methods and compositions of the invention. **Fig. 3A** depicts the method of selecting a positive cell transfectant from a first cell pool cultured with 1 µg/mL tunicamycin (Tn). Subsequently, a second cell culture with an increased concentration of tunicamycin, *e.g.* 2.5 µg/mL or 5 µg/mL, to enhance protein expression. **Fig. 3B**: depicts a method of selecting a positive cell transfectant from a first cell pool cultured with 1 µg/mL tunicamycin (Tn), and then serially increasing concentrations of Tn in subsequent cell cultures in order to optimize protein expression.

[0032] Figures 4A to 4B show FACS scatterplots representing various parameters of Hygromycin selectivity. Modified CHO cells comprise a YFP gene flanked by lox sites. Selection markers (antibiotic resistance gene and eGFP) flanked by lox sites incorporate at the YFP site and replace YFP via targeted integration with Cre recombinase. Random integrants express both YFP and eGFP **Fig. 4A**: Cells are transfected with a Cre recombinase vector and *hpt* expression vector comprising eGFP; but cultured without hygromycin in culture. **Fig. 4B**: Cells are transfected with a Cre recombinase vector and *hpt* expression vector comprising eGFP; in the presence of 400 µg/mL hygromycin.

[0033] Figures 5A to 5F show FACS scatterplots representing various parameters of Tunicamycin (Tn) selectivity. Modified CHO cells comprise a YFP gene flanked by lox sites. Selection markers (antibiotic resistance gene and eGFP) flanked by lox sites incorporate at the YFP site and replace YFP via targeted integration with Cre recombinase. Random integrants express both YFP and eGFP **Fig. 5A**: Cells are transfected with a Cre recombinase vector and CHO-GPT expression vector comprising eGFP; but without tunicamycin in culture. **Fig. 5B**: Cells are transfected with a Cre recombinase vector and CHO-GPT expression vector comprising eGFP; in the presence of 1 µg/mL Tn. **Fig. 5C**: Cells are transfected with a Cre recombinase vector and CHO-GPT expression vector comprising eGFP; in the presence of 2.5 µg/mL Tn. **Fig. 5D**: Cells are transfected with a Cre recombinase vector and Human GPT expression vector comprising eGFP; but without tunicamycin in culture. **Fig. 5E**: Cells are transfected with a Cre recombinase vector and Human GPT expression vector comprising

eGFP; in the presence of 1 µg/mL Tn. **Fig. 5F**: Cells are transfected with a Cre recombinase vector and Human GPT expression vector comprising eGFP; in the presence of 2.5 µg/mL Tn.

[0034] Figures 6A and 6B show GPT expressing cell pools compared to non-GPT expressing pools in their relative ability to enhance expression of an operably linked GOI, such as eGFP. **Fig. 6A**: illustrates the relative number of gene copies of CHO-GPT as measured by PCR for cell pools as follows: Pool-49 cells (no exogenous GPT added) without Tn selection; Pool-49 cells (no exogenous GPT) with 5 ug Tn selection; Pool-1 cells naturally express higher amounts of GPT (data not shown), and are tested without Tn selection; Pool-78 cells (no exogenous GPT) without Tn selection; CHO cells expressing exogenously-added *hpt* and 400 µg/mL Hygromycin selection; CHO cells expressing exogenous GPT under 1 µg/mL Tn selection conditions; CHO cells expressing exogenous GPT selected from a 1 µg/mL Tn selection pool further cultured in 1 µg/mL Tn; CHO cells expressing exogenous GPT selected from a 1 µg/mL Tn selection pool further cultured in 2.5 µg/mL Tn; CHO cells expressing exogenous GPT selected from a 1 µg/mL Tn selection pool further cultured in 5 µg/mL Tn. **Fig. 6B**: illustrates the relative number of gene copies of a gene of interest, eGFP, as measured by qPCR for the same cell pools (as Fig. 6A).

[0035] Figures 7A to 7D illustrate glycoform characteristics of Fc-fusion protein 1 (FcFP1) produced from cell culture as follows, **Fig. 7A:** CHO cells not expressing GPT using a standard protocol (Lot B10002M410), compared to **Fig. 7B:** CHO cells expressing CHO-GPT and no Tn selection (Lot 110728). **Fig. 7C:** CHO cells expressing CHO-GPT and selected with 1 µg/mL Tn (Lot 110728-01), compared to **Fig. 7D:** CHO cells expressing CHO-GPT and selected with 5 µg/mL Tn (Lot 110728-02). Each chromatogram indicates fractions containing sialylated residues as follows: 0SA = zero sialic acid residues; 1SA = one sialic acid residue; 2SA = two sialic acid residues; 3SA = three sialic acid residues; 4SA = four sialic acid residues.

[0036] Figure 8 illustrates the overlapping glycosylation profile of Fc-fusion protein 1 (FcFP1) sampled from (A) Lot B10002M410, (B) Lot 110728, (C) Lot 110728-01, and (D) Lot 110728-02. The glycoprofiles of each protein produced from the GPT lots are compatible with the reference standard protein and the major glycoform species are consistently produced. It is apparent that no new and unique species of glycoforms were produced in the GPT lots compared to the reference standard protein.

DETAILED DESCRIPTION

[0037] Before the present methods are described, it is to be understood that this invention is not limited to particular methods, and experimental conditions described, as such methods and conditions may 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.

[0038] As used in this specification and the appended claims, the singular forms “a”, “an”, and “the” include plural references unless the context clearly dictates otherwise. Thus for example, a

reference to “a method” includes one or more methods, and/or steps of the type described herein and/or which will become apparent to those persons skilled in the art upon reading this disclosure.

[0039] Unless defined otherwise, or otherwise specified, 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.

[0040] Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, particular methods and materials are now described. All publications mentioned herein are incorporated herein by reference in their entirety.

[0041] A variety of genes well-known in the art may confer a selectable phenotype on mammalian cells in culture. Commonly, selectable marker genes express proteins, usually enzymes that confer resistance to various antibiotics in cell culture. In some selective conditions, cells that express a flourescent protein marker are made visible, and are thus selectable. Examples in the art include beta-lactamase (bla; beta- lactam antibiotic resistance gene or ampR; ampicillin resistance gene), bls (blasticidin resistance acetyl transferase gene), hygromycin phosphotransferase (hpt; hygromycin resistance gene), and others.

[0042] The methods described herein rely on the use of tunicamycin and enzymes (markers) that can allow cells resistant to tunicamycin to grow in cell culture. Tunicamycin (Tn) is mixture of antibiotics that act as inhibitors of bacterial and eukaryote N-acetylglucosamine transferases preventing formation of N-acetylglucosamine lipid intermediates and glycosylation of newly synthesized glycoproteins. (King, I.A., and Tabiowo, A., 1981, Effect of tunicamycin on epidermal glycoprotein and glycosaminoglycan synthesis in vitro. Biochem. J., 198(2):331-338). Tn is cytotoxic because it specifically inhibits UDP-N-acetylglucosamine: dolichol phosphate N-acetylglucosamine-1-P transferase (GPT), an enzyme that catalyzes the initial step of the biosynthesis of dolichol-linked oligosaccharides. In the presence of tunicamycin, asparagine-linked glycoproteins made in the endoplasmic reticulum (ER) are not glycosylated with N-linked glycans, and therefore may not fold correctly in the ER and thus, may be targeted for breakdown (Koizumi, et al. 1999, Plant Physiol. 121(2):353–362). Hence, Tn is a notable inducer of the unfolded protein response (UPR) which leads to apoptosis in bacterial and eukaryotic cells.

[0043] The gene for uridine diphosphate GPT (also known as GlcNAc-1-P transferase) was identified as being overexpressed under certain cellular conditions in order to confer resistance to Tn (Criscuolo and Krag, 1982, *J Biol Chem*, 263(36):19796-19803; Koizumi, et al., 1999, *Plant Physiology*, Vol. 121, pp. 353–361). The gene encoding GPT, also described as GenBank Accn. No. M36899 (SEQ ID NO: 2), was isolated from a Tn-resistant Chinese hamster ovary cell line and encodes a 408 amino acid protein (SEQ ID NO: 3) (Scocca and Krag, 1990, *J Biol Chem* 265(33):20621-20626; Lehrman, M. et al., 1988, *J Biol Chem* 263(36):19796-803). Hamster GPT was overexpressed in yeast cells (*S. pombe*) and conferred Tn resistance in these cells; also providing a convenient source for the purification of the GPT enzyme (Scocca

JR, et al. 1995, *Glycobiology*, 5(1):129-36). Transcript levels of GPT were analyzed in hybridoma cells (B cells expressing IgG, vs. quiescent B cells) whereas it was observed that IgG-producing cells did not exhibit increased levels of GPT transcript or activity, yet a small increase in GPT was seen in the transition from quiescent to active B cells. It was concluded that GPT levels may correspond with the early development of proliferative response to LPS (antigen) stimulation in B cells (Crick, D.C. et al, 1994, *J Biol Chem* 269(14):10559-65).

[0044] Furthermore, it was previously unknown whether altering the expression of GPT, with or without the presence of Tn, in a cellular expression system will have an effect on the glycosylation of protein product, and therefore on product quality. It is understood that optimal and consistent glycosylation is a critical protein attribute in the production of therapeutic glycoproteins.

[0045] The present invention provides an improved method for production of recombinant proteins in mammalian cell systems utilizing a mammalian Tn-resistance gene, GPT, as a regulatable selection marker, whereas increased copy number of a gene of interest operably linked to GPT correlates with increased random integration of a GPT expression cassette into the cell.

[0046] The art has recognized that the manufacture of therapeutic proteins, particularly glycoproteins, relies on mammalian-type expression systems that mimic natural glycosylation of such proteins. (For review, see Bork, K. et al, 2009, J Pharm Sci. 98(10):3499-3508.) For example, the terminal monosaccharide of certain glycoproteins such as N-linked complex glycans is typically occupied by sialic acid. Sialylation may affect the glycoprotein's pharmacokinetic properties, such as absorption, serum half-life, and clearance, or other physicochemical or immunogenic properties of the glycoprotein. Overexpressed recombinant glycoproteins often have incomplete or inconsistent glycosylation. Reliable methods are critical for process consistency and quality of therapeutic glycoproteins produced in mammalian cell lines.

[0047] The present invention also provides an improved method for the glycosylation of recombinant proteins, *i.e*. a method for making glycoproteins, in mammalian cell systems in order to provide consistent quality yield of the desired proteins.

Definitions

[0048] DNA regions are operably linked when they are functionally related to each other. For example, a promoter is operably linked to a coding sequence if the promoter is capable of participating in the transcription of the sequence; a ribosome-binding site is operably linked to a coding sequence if it is positioned so as to permit translation. Generally, operably linked can include, but does not require, contiguity. In the case of sequences such as secretory leaders, contiguity and proper placement in a reading frame are typical features. A production enhancing sequence, such as a promoter, is operably linked to a gene of interest (GOI) where it is

functionally related to the GOI, for example, where its presence results in increased expression of the GOI.

[0049] As such, the phrase “operably linked”, such as in the context of DNA expression vector constructs, a control sequence, *e.g.*, a promoter or operator or marker, is appropriately placed at a position relative to a coding sequence such that the control sequence directs or permits the production of a polypeptide/protein of interest encoded by the coding sequence. For example, where a selection marker is required for cells to survive in certain culture conditions, the gene of interest is operably linked to the selection marker gene because expression will not occur without the presence of an operable selection marker protein.

[0050] "Promoter" as used herein indicates a DNA sequence sufficient to direct transcription of a DNA sequence to which it is operably linked, *i.e.,* linked in such a way as to permit transcription of the gene of interest and/or selection marker gene when the appropriate signals are present. The expression of a gene may be placed under control of any promoter or enhancer element known in the art.

[0051] An “expression vector” in the context of the present invention may be any suitable vector, including chromosomal, non-chromosomal, and synthetic nucleic acid vectors (a nucleic acid sequence comprising a suitable set of expression control elements). Examples of such vectors include derivatives of SV40, bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, and viral nucleic acid (RNA or DNA) vectors. In one embodiment, an Fc-fusion protein or polypeptide-encoding nucleic acid molecule is comprised in a naked DNA or RNA vector, including, for example, a linear expression element (as described in, for instance, Sykes and Johnston, 1997, *Nat Biotech* 12, 355-59), a compacted nucleic acid vector (as described in for instance US6,077,835 and/or WO00/70087), or a plasmid vector such as pBR322, pUC 19/18, or pUC 118/119. Such nucleic acid vectors and the usage thereof are well known in the art (see, for instance, US5,589,466 and US5,973,972).

[0052] As used herein "operator" indicates a DNA sequence that is introduced in or near a gene in such a way that the gene may be regulated by the binding of a repressor protein to the operator and, as a result, prevent or allow transcription of the GOI, *i.e*. a nucleotide encoding a polypeptide or protein of interest.

[0053] Ribosome binding sites include “internal ribosome entry sites” (IRESs) or may include a 5’ cap. Many IRES sequences are well-known in the art. IRES represents a translation control sequence, wherein the IRES site is typically located 5’ of a gene of interest and allows translation of the RNA in a cap-independent manner. Transcribed IRESs may directly bind ribosomal subunits such that the location of the mRNA’s initiator codons is oriented properly in the ribosome for translation. IRES sequences are typically located in the 5’ UTR of the mRNA (directly upstream of the initiation codon). IRESs functionally replace the need for various protein factors that interact with eukaryotic translation machinery.

[0054] The terms “enhanced” or “improved” when used to describe protein expression include

an increase in the quantity and/or consistency of quality of the protein (*i.e.* gene product) produced by the expression system or methods of the invention. As such, this includes an enhancement of at least about 1.5-fold to at least about 3-fold enhancement in expression over what is typically observed by random integration into a genome, for example, as compared to a pool of integrants using another selectable marker construct. As such, fold-expression enhancement observed for proteins of interest is compared to an expression level of the same gene, measured under substantially the same conditions, in the absence of an expression cassette or cell of the invention comprising a GPT gene, or in the presence of an expression cassette or cell comprising a different selectable marker. Expression enhancement may also be measured by the resulting number of random integration events. Enhanced recombination efficiency includes an enhancement of the ability of a locus to recombine (for example, employing recombinase-recognition sites). Enhancement refers to a measurable efficiency over random recombination, which is typically 0.1%. In certain conditions, enhanced recombination efficiency is about 10-fold over random, or about 1%. Unless specified, the claimed invention is not limited to a specific recombination efficiency. Expression enhancement may also be measured by the resulting number of gene copies as measured by quantitative polymerase chain reaction (qPCR), or other well-known technique.

[0055] Enhanced or improved product also refers to the more consistent quality, for example, post-translational modifications observed with the GPT expression system of the invention. Consistent quality includes having *e.g.* a desirable glycosylation profile after replicate production lines. Consistency, with respect to quality, refers to a degree of uniformity and standardization, whereas replicate production batches are essentially free from variation. Calculating a Z-number to measure consistency is taught herein. Other statistical measures are known in the art for measuring consistency.

[0056] The phrase “selective pressure” is the force or stimulus applied to a living organism (*e.g.* a cell) or system (*e.g.* as an expression system) which alters the behavior and survival (such as ability to survive) of the living organism or system within a given environment.

[0057] The phrase “gene amplification” means an increase in the number of identical copies of a gene sequence. Certain cellular processes are characterized by the production of multiple copies of a particular gene or genes that amplify the phenotype that the gene confers on the cell, for example antibiotic resistance.

[0058] Where the phrase “exogenously added gene” or “exogenously added GOI” is employed with reference to an expression cassette, the phrase refers to any gene not present within the cell genome as found in nature, or an additional gene copy integrated into (a different locus within) the genome. For example, an “exogenously added gene” within a CHO genome (*e.g.,* an selectable marker gene), can be a hamster gene not found within the particular CHO locus in nature (*i.e.,* a hamster gene from another locus in the hamster genome), a gene from any other species (*e.g.*, a human gene), a chimeric gene (*e.g.,* human/mouse), or can be a hamster gene not found within the CHO genome in nature (*i.e.,* a hamster gene having less than 99.9%

identity to the gene from another locus in the hamster genome), or any other gene not found in nature to exist within the CHO natural genome.

[0059] Random integration events differ from targeted integration events, whereas insertion of a gene into the genome of the cell is not site-specific in random integration events. An example of targeted integration is homologous recombination. Random (nonhomologous) integration means that the location (locus) of the resulting integrant is not known or specified. Random integration is thought to occur by nonhomologous end joining (NHEJ), however is not limited to this method.

[0060] Selection efficiency means the percent population of surviving cells expressing the selectable marker and, if applicable, the protein of interest under the control of the selectable marker.

[0061] Percent identity, when describing a Tn-resistance protein, is meant to include homologous sequences that display the recited identity along regions of contiguous homology, but the presence of gaps, deletions, or insertions that have no homolog in the compared sequence are not taken into account in calculating percent identity. In explaining the usage of "percent identity" in this context, the following amino acid sequence comparison will be referred to:

```
1   MWAFPELPLPLPLLVNLIGSLLGFVATVTLIPAFRSHFIAARLCGQDLNKLSQQQIPESQ   60  GPT_MOUSE
1   MWAFPELPL--PLLVNLFGSLLGFVATVTLIPAFRSHFIAARLCGQDLNKLSRQQIPESQ   58  GPT_CRIG
```

[0062] As used herein, a "percent identity" determination between the "GPT_CRIG" sequence above (for a Chinese hamster GPT) with a mouse homolog ("GPT_MOUSE") would not include a comparison of hamster amino acids 10 and 11, since the hamster homolog has no homologous sequence to compare in an alignment (*i.e.,* the mouse GPT has an insertion at that point, or the hamster homolog has a gap or deletion, as the case may be). Thus, in the comparison above, the percent identity comparison would extend from the "MWA" at the 5' end to the "ESQ" at the 3' end. In that event, the mouse homolog differs only in that it has an "R" at hamster GPT position 51. Since the comparison is over 58 contiguous bases in a 60 base pair stretch, with only one amino acid difference (which is not a gap, deletion, or insertion), there is over 98% identity between the two sequences (hamster and mouse) from hamster GPT position 1 to hamster GPT position 58 (because "percent identity" does not include penalties for gaps, deletions, and insertions). Although the above example is based on an amino acid sequence, it is understood that nucleic acid sequence percent identity would be calculated in the same manner.

[0063] The term "cell" includes any cell that is suitable for expressing a recombinant nucleic acid sequence. Cells include those of prokaryotes and eukaryotes (single-cell or multiple-cell), bacterial cells (e.g., strains of *E. coli*, *Bacillus* spp., *Streptomyces* spp., etc.), mycobacteria cells, fungal cells, yeast cells (e.g. *S. cerevisiae*, *S. pombe*, *P. partoris*, *P. methanolica*, etc.), plant cells, insect cells (e.g. SF-9, SF-21, baculovirus-infected insect cells, *Trichoplusia ni*, etc.), non-

human animal cells, mammalian cells, human cells, or cell fusions such as, for example, hybridomas or quadromas. In certain embodiments, the cell is a human, monkey, ape, hamster, rat or mouse cell. In other embodiments, the cell is eukaryotic and is selected from the following cells: CHO (e.g. CHO K1, DXB-11 CHO, Veggie-CHO), COS (e.g. COS-7), retinal cells, Vero, CV1, kidney (e.g. HEK293, 293 EBNA, MSR 293, MDCK, HaK, BHK21), HeLa, HepG2, WI38, MRC 5, Colo25, HB 8065, HL-60, Jurkat, Daudi, A431 (epidermal), CV-1, U937, 3T3, L cell, C127 cell, SP2/0, NS-0, MMT cell, tumor cell, and a cell line derived from an aforementioned cell. In some embodiments, the cell comprises one or more viral genes, *e.g.* a retinal cell that expresses a viral gene (e.g. a PER.C6® cell).

[0064] The phrase "integrated cell density", or "ICD" means the density of cells in a culture medium taken as an integral over a period of time, expressed as cell-days per mL. In some embodiments, the ICD is measured around the twelfth day of cells in culture.

[0065] "Glycosylation" or the phrase "glycosylating a protein" includes the formation of glycoproteins whereas oligosaccharides are attached either to the side chain of the asparagine (Asn) residue (*i.e. N*-linked) or serine (Ser)/threonine (Thr) residue (*i.e. O*-linked) of a protein. Glycans can be homo- or heteropolymers of monosaccharide residues, which can be linear or branched. *N*-linked glycosylation is known to initiate primarily in the endoplasmic reticulum, whereas *O*-linked glycosylation is shown to initiate in either the ER or Golgi apparatus.

[0066] An "*N*-glycan protein" or an "*N*-glycan protein substrate" includes proteins that contain or can accept *N*-linked oligosaccharides. *N*-glycans can be composed of N-acetyl galactosamine (GalNAc), mannose (Man), fucose (Fuc), galactose (Gal), neuraminic acid (NANA), and other monosaccharides, however *N*-glycans usually have a common core pentasaccharide structure including: three mannose and two N-acetylglucosamine (GlcNAc) sugars. Proteins with the consecutive amino acid sequence (*i.e.* sequon) Asn-X-Ser or Asn-X-Thr, where X is any amino acid except proline, can provide an attachment site for *N*-glycans.

General Description

[0067] The invention is based at least in part on the discovery that under certain conditions recombinant proteins may be produced in a cell wherein the gene encoding the protein is operably linked to a Tn-resistance gene, GPT, and selection of a protein-producing cell is formatted to increase random integration events in the cell genome and thus increase copy number of the gene of interest, and ultimately protein production.

[0068] The invention is also based at least in part on the finding that the protein-producing cell may be optimized to express proteins with consistent and reliable post-translational modifications. GPT expression cassettes can also be integrated in a cellular genome, as in expression constructs, such as via expression vectors, using various gene editing techniques known in the art. Expression vectors comprising GPT can be integrated into a genome by random or targeted recombination such as, homologous recombination or recombination mediated by recombinases that recognize specific recombination sites (*e.g.*, Cre-lox-mediated

recombination).

[0069] Homologous recombination in eukaryotic cells can be facilitated by introducing a break in the chromosomal DNA at the integration site. Model systems have demonstrated that the frequency of homologous recombination during gene targeting increases if a double-strand break is introduced within the chromosomal target sequence. This may be accomplished by targeting certain nucleases to the specific site of integration. DNA-binding proteins that recognize DNA sequences at the target locus are known in the art. Gene targeting vectors are also employed to facilitate homologous recombination. In the absence of a gene targeting vector for homology directed repair, the cells frequently close the double-strand break by non-homologous end-joining (NHEJ) which may lead to deletion or insertion of multiple nucleotides at the cleavage site. Gene targeting vector construction and nuclease selection are within the skill of the artisan to whom this invention pertains.

[0070] In some examples, zinc finger nucleases (ZFNs), which have a modular structure and contain individual zinc finger domains, recognize a particular 3-nucleotide sequence in the target sequence (*e.g.* site of targeted integration). Some embodiments can utilize ZFNs with a combination of individual zinc finger domains targeting multiple target sequences.

[0071] Transcription activator-like (TAL) effector nucleases (TALENs) may also be employed for site-specific genome editing. TAL effector protein DNA-binding domain is typically utilized in combination with a non-specific cleavage domain of a restriction nuclease, such as FokI. In some embodiments, a fusion protein comprising a TAL effector protein DNA-binding domain and a restriction nuclease cleavage domain is employed to recognize and cleave DNA at a target sequence within the locus of the invention (Boch J et al., 2009 *Science* 326:1509-1512).

[0072] RNA-guided endonucleases (RGENs) are programmable genome engineering tools that were developed from bacterial adaptive immune machinery. In this system—the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) immune response—the protein Cas9 forms a sequence-specific endonuclease when complexed with two RNAs, one of which guides target selection. RGENs consist of components (Cas9 and tracrRNA) and a target-specific CRISPR RNA (crRNA). Both the efficiency of DNA target cleavage and the location of the cleavage sites vary based on the position of a protospacer adjacent motif (PAM), an additional requirement for target recognition (Chen, H. et al, *J. Biol. Chem*. published online March 14, 2014, as Manuscript M113.539726).

[0073] Still other methods of homologous recombination are available to the skilled artisan, such as BuD-derived nucleases (BuDNs) with precise DNA-binding specificities (Stella, S. et al. *Acta Cryst*. 2014, D70, 2042-2052). Precise genome modification methods are chosen based on the tools available compatible with unique target sequences within the genome so that disruption of the cell phenotype is avoided.

[0074] Cells and methods are provided for stably integrating a nucleic acid sequence (gene of interest) into a mammalian cell, wherein the nucleic acid sequence is capable of enhanced expression by virtue of being integrated with a GPT sequence. Compositions and methods are

also provided for using GPT in connection with expression constructs, for example, expression vectors, and for adding an exogenous GPT into a mammalian cell of interest. Cells and methods are provided for use in a consistent yet robust method of making glycoproteins, particularly therapeutic glycoproteins.

Construction of a GPT selection marker cassette

[0075] Expression vectors comprising an operative GPT expression cassette are provided herein. The expression cassette comprises the necessary regulatory elements to permit and drive transcription and translation of mammalian GPT and the desired gene product.

[0076] Various combinations of the genes and regulatory sequences described herein can also be developed. Examples of other combinations of the appropriate sequences described herein that can also be developed include sequences that include multiple copies of the GPT genes disclosed herein, or sequences derived by combining the disclosed GPT with other nucleotide sequences to achieve optimal combinations of regulatory elements. Such combinations can be contiguously linked or arranged to provide optimal spacing of GPT oriented to the gene of interest and the regulatory elements.

[0077] Homologous sequences of genes encoding GPT are known to exist in cells from other mammalian species (such as, for example, humans; see Fig. 2) as well as in cell lines derived from other mammalian tissue types, and can be isolated by techniques that are well-known in the art. An exemplary list of mammalian GPT amino acid sequences is provided in Figure 2. Changes in nucleotide sequence, such as codon optimization, can be made to nucleotide sequences set forth in SEQ ID NOs:2 and 11-17 in order to permit optimal expression of the corresponding GPT proteins set forth in SEQ ID NOs:3-10. In addition, changes can be made in the amino acid sequence set forth in SEQ ID NOs:3-10 by making changes to the nucleotide sequences encoding GPT. Such techniques including, but not limited to, site-directed or random mutagenesis techniques are well known in the art.

[0078] The resulting GPT variants can then be tested for GPT activity as described herein, *e.g.* tested for resistance to tunicamycin. GPT proteins that are at least about 93% identical, or at least about 95% identical, or at least about 96% identical, or at least about 97% identical, or at least about 98% identical in amino acid sequence to SEQ ID NO:3 having GPT activity are isolatable by routine experimentation, and are expected to exhibit the same resistance to Tn, selectivity efficiency and post-translational benefits as for SEQ ID NO:3. Accordingly, mammalian homologs of GPT and variants of GPT are also encompassed by embodiments of the invention. Figs. 2A to 2C show an alignment of various mammalian GPT amino acid sequences (namely, SEQ ID NOs: 3-10). The mammalian GPT sequences (nucleic acid and amino acid) are conserved among hamster, human, mouse and rat genomes. Table 1 identifies exemplary mammalian GPT proteins and their degree of homology.

TABLE 1A: Amino acid identity of GPT homologs

Animal	SEQ ID NO	%id Human	%id Mouse	%id Rat	%id Hamster
Hamster	3	93.87	96.08	96.08	-
Mouse	10	94.12	-	97.07	96.08
Human	4	-	94.12	93.63	93.87
Rat	9	93.63	97.07	-	96.08

TABLE 1B: Nucleic acid identity of representative GPT homologs

Animal	SEQ ID NO	%id Hamster
Hamster	2	-
Mouse	11	92
Human	12	92
Rat	13	94
Macaque	14	92
Chimp	15	92

[0079] Cell populations expressing enhanced levels of a protein of interest can be developed using the GPT/tunicamycin methods provided herein. The absolute level of expression will vary with the specific protein, depending on how efficiently the protein is processed by the cell.

[0080] Accordingly, the invention also includes a GPT-expressing nucleotide sequence selected from the group consisting of SEQ ID NOs:2 and 11-17. The invention also encompasses a GPT-expressing nucleotide sequence that is at least 92% identical, at least 93% identical, at least 94% identical, at least 95% identical, at least 96% identical, at least 98% identical, or at least 99% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs:2 and 11-17.

[0081] The invention includes vectors comprising SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:12. Vectors comprising a mammalian GPT gene, and optional regulatory elements, include vectors for transient or stable transfection.

[0082] In one embodiment, the GPT gene is employed to enhance the expression of a GOI, as illustrated in Fig. 1. Fig. 1 shows a GOI operably linked with an IRES sequence and a GPT selectable marker. The GPT cassette further includes a promoter sequence, e.g. SV40 promoter, and a polyadenylation (poly(A)) sequence, e.g. SV40 poly(A).

[0083] The expression-enhancing cassette (including GPT and an upstream promoter) is optimally integrated in a cell genome. Using the methods of the invention, a GOI is expressed within the GPT expression cassette under culture conditions based on increasing concentrations of Tn (Fig. 3A or Fig. 3B). A FACS readout, such as that shown in Figures 5B, 5C, 5E and 5F, exemplifies the distribution of expression in a stably transfected population of cells, in particular the dramatic increase in selection efficiency using mammalian Tn-resistant selection markers,

CHO-GPT and hGPT. Mammalian GPT expression further enhances expression of a gene product of interest, for example production of a fluorescent protein, eGFP. Consecutive cultures of increasing concentrations of Tn result in an enhanced expression of about two-fold in comparison to the GOI expressed in an expression system using GPT under culture conditions based on one concentration of Tn, such as that exemplified in Fig. 6B.

[0084] The invention includes a mammalian cell comprising such a GPT gene wherein the GPT gene is exogenous and is integrated into the cell genome by the methods of the invention. Cells comprising such a GPT gene having at least one exogenously-added gene of interest (GOI) that is upstream or downstream to the GPT gene.

[0085] In various embodiments, expression of a GOI can be enhanced by placing the GOI under the control of a mammalian selectable marker GPT. In other embodiments, the random integration events of a GOI can be enhanced by placing the GOI under the control of a mammalian selectable marker GPT and providing cell culture conditions comprising greater than 0.5 µg/mL Tn concentration. In some embodiments, the cell culture conditions comprise greater than 1 µg/mL Tn concentration. A regulatory element may be operably linked to the GOI where expression of the GOI—at the selected distance from the GOI and GPT (in the 5' or 3' direction)—retains the ability to enhance expression of the GOI over, for example, expression typically observed due to a random integration event. In various embodiments, enhancement is at least about 1.5-fold to about 2-fold or more. Enhancement in expression as compared to a random integration, or random expression, is about 1.5-fold or about 2-fold or more.

[0086] In another embodiment, uniformly glycosylated proteins can be attained using the methods and compositions of the invention. As shown in Table 4, GPT/GOI recombinant protein batches treated with Tn allow replicate batches with equivalent glycosylation profiles. As such, enhanced protein expression such as consistent glycosylation profiles can be directly compared by calculating Z-number as taught herein. The Z-number equation takes into consideration takes into account the relative number of peaks on a chromatogram representing sialic acid (SA) moieties, as well as the relative shape and intensity of each peak. Z-number is based on the area occupied by each peak and may be used as a measure of consistency for complex glycoproteins (see *e.g.* Figs. 7A-7D, Fig 8, and Example 3, described herein.

[0087] Protein expression optimization can also be achieved for each GOI, including, for example, expression cassette orientation or codon optimization. Protein optimization may also be achieved by varying the incremental Tn concentration in the cell culture methods.

[0088] Recombinant expression vectors can comprise synthetic or cDNA-derived DNA fragments encoding a protein, operably linked to a suitable transcriptional and/or translational regulatory element derived from mammalian, viral or insect genes. Such regulatory elements include transcriptional promoters, enhancers, sequences encoding suitable mRNA ribosomal binding sites, and sequences that control the termination of transcription and translation, as described in detail herein. Mammalian expression vectors can also comprise nontranscribed elements such as an origin of replication, other 5' or 3' flanking nontranscribed sequences, and

5' or 3' nontranslated sequences such as splice donor and acceptor sites. Additional selectable marker genes (such as fluorescent markers) to facilitate recognition of transfectants may also be incorporated.

[0089] In another embodiment, the vector comprises a nucleic acid molecule (or gene of interest) encoding a protein of interest, including an expression vector comprising the nucleic acid molecules (genes) described wherein the nucleic acid molecule (gene) is operably linked to an expression control sequence.

[0090] A vector comprising a gene of interest (GOI) is provided, wherein the GOI is operably linked to an expression control sequence suitable for expression in a mammalian host cell.

[0091] Useful promoters that may be used in the invention include, but are not limited to, the SV40 early promoter region, the promoter contained in the 3' long terminal repeat of Rous sarcoma virus, the regulatory sequences of the metallothionein gene, mouse or human cytomegalovirus IE promoter (Gossen et al., (1995) Proc. Nat. Acad. Sci. USA 89:5547-5551), the cauliflower mosaic virus 35S RNA promoter, and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase, promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I; insulin; immunoglobulin; mouse mammary tumor virus; albumin; α-fetoprotein; α1-antitrypsin; β-globin; and myosin light chain-2.

[0092] Nucleic acid molecules of the invention may also be operably linked to an effective poly (A) termination sequence, e.g. SV40 poly(A), an origin of replication for plasmid product in E. coli, and/or a convenient cloning site (e.g., a polylinker). Nucleic acids may also comprise a regulatable inducible promoter (inducible, repressable, developmentally regulated) as opposed to a constitutive promoter such as CMV IE (the skilled artisan will recognize that such terms are actually descriptors of a degree of gene expression under certain conditions).

[0093] The invention provides methods for producing a protein of interest whereas an expression vector is provided comprising a gene of interest (GOI) is provided. Such expression vectors may be used for recombinant production of any protein of interest. Transcriptional and translational control sequences in expression vectors useful for transfecting vertebrate cells may be provided by viral sources. For example, commonly used promoters and enhancers are derived from viruses such as polyoma, adenovirus 2, simian virus 40 (SV40), and human cytomegalovirus (CMV). Viral genomic promoters, control and/or signal sequences may be utilized to drive expression, provided such control sequences are compatible with the host cell chosen. Non-viral cellular promoters can also be used (*e.g.,* the β-globin and the EF-1α promoters), depending on the cell type in which the recombinant protein is to be expressed.

[0094] DNA sequences derived from the SV40 viral genome, for example, the SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites may be used to provide other

genetic elements useful for expression of a heterologous DNA sequence. Early and late promoters are particularly useful because both are obtained easily from the SV40 virus as a fragment that also comprises the SV40 viral origin of replication (Fiers *et al.,* Nature 273:113, 1978). Smaller or larger SV40 fragments may also be used. Typically, the approximately 250 bp sequence extending from the Hind III site toward the BglI site located in the SV40 origin of replication is included.

[0095] Bicistronic expression vectors used for the expression of multiple transcripts have been described previously (Kim S. K. and Wold B. J., Cell 42:129, 1985; Kaufman *et al.* 1991, supra) and can be used in combination with a GPT expression system. Other types of expression vectors will also be useful, for example, those described in U.S. Pat. No. 4,634,665 (Axel *et al.*) and U.S. Pat. No. 4,656,134 (Ringold *et al.*).

[0096] An integration site, for example, a recombinase recognition site, can be placed 5' or 3' to the gene sequence encoding the POI. One example of a suitable integration site is a lox p site. Another example of a suitable integration site is two recombinase recognition sites, for example, selected from the group consisting of a lox p site, lox and a lox 5511 site.

Gene Amplification Cassettes and Expression Vectors Thereof

[0097] Useful regulatory elements, described previously or known in the art, can also be included in the nucleic acid constructs used to transfect mammalian cells. Figure 1 exemplifies an operative cassette in a GPT vector further comprising a promoter sequence, IRES sequence, gene of interest, and poly(A) sequence.

[0098] An expression vector in the context of the present invention may be any suitable vector, including chromosomal, non-chromosomal, and synthetic nucleic acid vectors (a nucleic acid sequence comprising a suitable set of expression control elements). Examples of such vectors include derivatives of SV40, bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, and viral nucleic acid (RNA or DNA) vectors. In one embodiment, an antibody-encoding nucleic acid molecule is comprised in a naked DNA or RNA vector, including, for example, a linear expression element (as described in, for instance, Sykes and Johnston, *Nat Biotech* 12, 355-59 (1997)), a compacted nucleic acid vector (as described in for instance US 6,077,835 and/or WO 00/70087), or a plasmid vector such as pBR322, pUC 19/18, or pUC 118/119. Such nucleic acid vectors and the usage thereof are well known in the art (see, for instance, US 5,589,466 and US 5,973,972).

[0099] An expression vector may alternatively be a vector suitable for expression in a yeast system. Any vector suitable for expression in a yeast system may be employed. Suitable vectors include, for example, vectors comprising constitutive or inducible promoters such as yeast alpha factor, alcohol oxidase and PGH (reviewed in: F. Ausubel et al., ed. Current Protocols in Molecular Biology, Greene Publishing and Wiley InterScience New York (1987), and Grant et al., *Methods in Enzymol* 153, 516-544 (1987)).

[00100] In certain embodiments, the vector comprises a nucleic acid molecule (or gene of

interest) encoding a protein of interest, including an expression vector comprising the nucleic acid molecules (genes) described wherein the nucleic acid molecule (gene) is operably linked to an expression control sequence suitable for expression in the host cell.

[00101] Expression control sequences are engineered to control and drive the transcription of genes of interest, and subsequent expression of proteins in various cell systems. Plasmids combine an expressible gene of interest with expression control sequences (i.e. expression cassettes) that comprise desirable regulatory elements such as, for example, promoters, enhancers, selectable markers, operators, etc. In an expression vector of the invention, GPT and the proteins of interest, such as antibody-encoding nucleic acid molecules, may comprise or be associated with any suitable promoter, enhancer, operator, repressor protein, poly (A) termination sequences and other expression-facilitating elements.

[00102] The expression of a gene of interest, such as an antibody-encoding nucleotide sequence, may be placed under control of any promoter or enhancer element known in the art. Examples of such elements include strong expression promoters (e. g., human CMV IE promoter/enhancer or CMV major IE (CMV-MIE) promoter, as well as RSV, SV40 late promoter, SL3-3, MMTV, ubiquitin (Ubi), ubiquitin C (UbC), and HIV LTR promoters).

[00103] In some embodiments, the vector comprises a promoter selected from the group consisting of SV40, CMV, CMV-IE, CMV-MIE, RSV, SL3-3, MMTV, Ubi, UbC and HIV LTR.

[00104] Nucleic acid molecules of the invention may also be operably linked to an effective poly (A) termination sequence, an origin of replication for plasmid product in E. coli, an antibiotic resistance gene as selectable marker, and/or a convenient cloning site (*e.g.*, a polylinker). Nucleic acids may also comprise a regulatable inducible promoter (inducible, repressable, developmentally regulated) as opposed to a constitutive promoter such as CMV IE (the skilled artisan will recognize that such terms are actually descriptors of a degree of gene expression under certain conditions).

[00105] Selectable markers are elements well-known in the art. In some circumstances, additional selectable markers may be employed, in addition to GPT, wherein such markers make the cells visible. Positive or negative selection may be used.

[00106] In some embodiments, the vector comprises one or more selectable marker genes encoding green fluorescent protein (GFP), enhanced green fluorescent protein (eGFP), cyano fluorescent protein (CFP), enhanced cyano fluorescent protein (eCFP), yellow fluorescent protein (YFP) or enhanced yellow fluorescent protein (eYFP).

[00107] For the purposes of this invention, gene expression in eukaryotic cells may be tightly regulated using a strong promoter that is controlled by an operator that is in turn regulated by a regulatory fusion protein (RFP). The RFP consists essentially of a transcription blocking domain, and a ligand-binding domain that regulates its activity. Examples of such expression systems are described in US20090162901A1, which is herein incorporated by reference in its entirety.

[00108] A number of operators in prokaryotic cells and bacteriophage have been well

characterized (Neidhardt, ed. Escherichia coli and Salmonella; Cellular and Molecular Biology 2d. Vol 2 ASM Press, Washington D.C. 1996). These include, but are not limited to, the operator region of the *LexA* gene of *E. coli*, which binds the LexA peptide, and the lactose and tryptophan operators, which bind the repressor proteins encoded by the *LacI* and *trpR* genes of *E. coli*. These also include the bacteriophage operators from the lambda P_R and the phage P22 *ant/mnt* genes which bind the repressor proteins encoded by lambda cI and P22 arc. In some embodiments, when the transcription blocking domain of the repressor protein is a restriction enzyme, such as NotI, the operator is the recognition sequence for that enzyme. One skilled in the art will recognize that the operator must be located adjacent to, or 3' to the promoter such that it is capable of controlling transcription by the promoter. For example, U.S. Pat. No. 5,972,650, which is incorporated by reference herein, specifies that *tetO* sequences be within a specific distance from the TATA box. In specific embodiments, the operator is preferably placed immediately downstream of the promoter. In other embodiments, the operator is placed within 10 base pairs of the promoter.

[00109] In certain embodiments, the operator is selected from the group consisting of tet operator (tetO), NotI recognition sequence, LexA operator, lactose operator, tryptophan operator and Arc operator (AO). In some embodiments, the repressor protein is selected from the group consisting of TetR, LexA, LacI, TrpR, Arc, LambdaC1 and GAL4. In other embodiments, the transcription blocking domain is derived from a eukaryotic repressor protein, *e.g.* a repressor domain derived from GAL4.

[00110] In an exemplary cell expression system, cells are engineered to express the tetracycline repressor protein (TetR) and a protein of interest is placed under transcriptional control of a promoter whose activity is regulated by TetR. Two tandem TetR operators (tetO) are placed immediately downstream of a CMV-MIE promoter/enhancer in the vector. Transcription of the gene encoding the protein of interest directed by the CMV-MIE promoter in such vector may be blocked by TetR in the absence of tetracycline or some other suitable inducer (*e.g.* doxycycline). In the presence of an inducer, TetR protein is incapable of binding tetO, hence transcription then translation (expression) of the protein of interest occurs. (See, *e.g.*, US Patent No. 7,435,553, which is herein incorporated by reference in its entirety.)

[00111] Another exemplary cell expression system includes regulatory fusion proteins such as TetR-ER_{LBD}T2 fusion protein, in which the transcription blocking domain of the fusion protein is TetR and the ligand-binding domain is the estrogen receptor ligand-binding domain (ER_{LBD}) with T2 mutations (ER_{LBD}T2; Feil et al. (1997) Biochem. Biophys. Res. Commun. 237:752-757). When *tetO* sequences were placed downstream and proximal to the strong CMV-MIE promoter, transcription of the nucleotide sequence of interest from the CMV-MIE/*tetO* promoter was blocked in the presence of tamoxifen and unblocked by removal of tamoxifen. In another example, use of the fusion protein Arc2-ER_{LBD}T2, a fusion protein consisting of a single chain dimer consisting of two Arc proteins connected by a 15 amino acid linker and the ER_{LBD}T2 (*supra*), involves an Arc operator (AO), more specifically two tandem arc operators immediately

downstream of the CMV-MIE promoter/enhancer. Cell lines may be regulated by Arc2-ER_{LBD}T2, wherein cells expressing the protein of interest are driven by a CMV-MIE/ArcO2 promoter and are inducible with the removal of tamoxifen. (See, *e.g.*, US 20090162901A1, which is herein incorporated by reference.)

[00112] In some embodiments, a vector of the invention comprises a CMV-MIE/TetO or CMV-MIE/AO2 hybrid promoter.

[00113] The vectors of the invention may also employ Cre-*lox* tools for recombination technology in order to facilitate the replication of a gene of interest. A Cre-*lox* strategy requires at least two components: 1) Cre recombinase, an enzyme that catalyzes recombination between two loxP sites; and 2) *lox*P sites (*e.g.* a specific 34-base pair bp sequence consisting of an 8-bp core sequence, where recombination takes place, and two flanking 13-bp inverted repeats) or mutant *lox* sites. (See, e.g. Araki et al. *PNAS* 92:160-4 (1995); Nagy, A. et al. *Genesis* 26:99-109 (2000); Araki et al. *Nuc Acids Res* 30(19):e103 (2002); and US20100291626A1, all of which are herein incorporated by reference). In another recombination strategy, yeast-derived FLP recombinase may be utilized with the consensus sequence FRT (see also, *e.g.* Dymecki, S. *PNAS* 93(12): 6191-6196 (1996)).

[00114] In another aspect, a gene (*i.e.* a nucleotide sequence encoding a recombinant polypeptide of the invention) is inserted upstream or downstream of the GPT gene of the expression cassette, and is optionally operably linked to a promoter, wherein the promoter-linked gene is flanked 5' by a first recombinase recognition site and 3' by a second recombinase recognition site. Such recombinase recognition sites allow Cre-mediated recombination in the host cell of the expression system. In some instances, a second promoter-linked gene is downstream (3') of the first gene and is flanked 3' by the second recombinase recognition site. In still other instances, a second promoter-linked gene is flanked 5' by the second recombinase site, and flanked 3' by a third recombinase recognition site. In some embodiments, the recombinase recognition sites are selected from a *lox*P site, a *lox*511 site, a *lox*2272 site, and a FRT site. In other embodiments, the recombinase recognition sites are different. In a further embodiment, the host cell comprises a gene capable of expressing a Cre recombinase.

[00115] In one embodiment, the vector comprises a first gene encoding a light chain of an antibody or a heavy chain of an antibody of the invention, and a second gene encoding a light chain of an antibody or a heavy chain of an antibody of the invention.

[00116] In some embodiments, the vector further comprises an X-box-binding-protein 1 (mXBP1) gene capable of further enhancing protein production/protein secretion through control of the expression of genes involved in protein folding in the endoplasmic reticulum (ER). (See, *e.g.* Ron D, and Walter P. *Nat Rev Mol Cell Biol*.8:519–529 (2007)).

[00117] Any cell is suitable for expressing a recombinant nucleic acid sequence of the invention. Cells used in the invention include mammalian cells such as non-human animal cells, human cells, or cell fusions such as, for example, hybridomas or quadromas. In certain embodiments, the cell is a human, monkey, hamster, rat or mouse cell. In other embodiments,

the cell is eukaryotic and is selected from the following cells: CHO (e.g. CHO K1, DXB-11 CHO, Veggie-CHO), COS (e.g. COS-7), retinal cells, Vero, CV1, kidney (e.g. HEK293, 293 EBNA, MSR 293, MDCK, HaK, BHK21), HeLa, HepG2, WI38, MRC 5, Colo25, HB 8065, HL-60, Jurkat, Daudi, A431 (epidermal), CV-1, U937, 3T3, L cell, C127 cell, SP2/0, NS-0, MMT cell, tumor cell, and a cell line derived from an aforementioned cell. In some embodiments, the cell comprises one or more viral genes, *e.g.* a retinal cell that expresses a viral gene (e.g. a PER.C6® cell).

[00118] In an even further aspect, the invention relates to a recombinant mammalian host cell, such as a transfectoma, which produces an immunoglobulin, such as an antibody or a bispecific molecule. Examples of such host cells include engineered mammalian cells such as CHO or HEK cells. For example, in one embodiment, the present invention provides a cell comprising a nucleic acid stably integrated into the cellular genome that comprises a sequence coding for expression of an antibody comprising a recombinant polypeptide of the present invention. In another embodiment, the present invention provides a cell comprising a non-integrated (i.e., episomal) nucleic acid, such as a plasmid, cosmid, phagemid, or linear expression element, which comprises a sequence coding for expression of an antibody comprising the recombinant polypeptide of the invention. In other embodiments, the present invention provides a cell line produced by stably transfecting a host cell with a plasmid comprising an expression vector of the invention.

[00119] Thus, in one aspect, the invention provides a cell containing (a) a recombinant polynucleotide that encodes an exogenously-added mammalian GPT gene and (b) a polynucleotide that encodes a multi-subunit protein. In some embodiments, the exogenously-added GPT gene is 90% identical to the nucleic acid sequence of SEQ ID NO: 2, non-limiting examples of which are provided in SEQ ID NOs:11-17, and the multi-subunit protein is an antibody. In other embodiments, the cell also contains an exogenously-added GPT gene, and regulatory elements. In one embodiment, the cell is a mammalian cell, such as a CHO cell used in the manufacture of biopharmaceuticals.

[00120] In another aspect, the invention provides a cell line derived from the cell described in the previous aspect. By “derived from”, what is meant is a population of cells clonally descended from an individual cell and having some select qualities, such as the ability to produce active protein at a given titer, or the ability to proliferate to a particular density. In some embodiments, the cell line, which is derived from a cell harboring the recombinant polynucleotide encoding a mammalian GPT gene and a polynucleotide encoding a multi-subunit protein, is capable of producing the multi-subunit protein at a titer of at least 3 grams per liter of media (g/L), at least 5 g/L, or at least 8 g/L. In some embodiments, the cell line can attain an integrated cell density (ICD) that is at least 30% greater, at least 50% greater, at least 60% greater, or at least 90% greater than the integrated cell density attainable by a cell line derived from what is essentially the same cell but without the recombinant polynucleotide encoding GPT.

[00121] A method for amplifying the GOI is provided. The exemplified methods apply

increasing concentrations of tunicamycin to a eukaryotic GPT expression system, thus amplifying the gene copy of a GOI operably linked to an exogenously-added mammalian GPT gene.

Proteins of Interest

[00122] A nucleic acid sequence encoding a protein of interest can be conveniently integrated into a cell comprising an Tn resistance marker gene and an IRES, and optionally flanked by recombinase recognition sites. Any protein of interest suitable for expression in mammalian cells can be used, however glycoproteins will especially benefit from the methods of the invention. For example, the protein of interest can be an antibody or antigen-binding fragment thereof, a bispecific antibody or fragment thereof, a chimeric antibody or fragment thereof, an ScFv or fragment thereof, an Fc-tagged protein (e.g. Trap protein) or fragment thereof, a growth factor or a fragment thereof, a cytokine or a fragment thereof, or an extracellular domain of a cell surface receptor or fragment thereof.

[00123] Glycoproteins with asparagine-linked (N-linked) glycans are ubiquitous in eukaryotic cells. Biosynthesis of these glycans and their transfer to polypeptides takes place in the endoplasmic reticulum (ER). N-glycan structures are further modified by a number of glycosidases and glycosyl-transferases in the ER and the Golgi complex. Protein production using the invention is directed at consistency of the native N-glycan structure in order to eliminate immunogenic epitopes (“glycotopes”).

[00124] Using the methods of the invention, recombinant protein lots display favorable characteristics. HPLC (with fluorescent detection) of replicate protein production batches demonstrated that the glycoproteins had uniform expression and glycosylation patterns, as exemplified in Figures 7-8 herein. A method of glycosylating a N-glycan protein substrate is provided, whereas a mammalian host cell encoding a nucleic acid molecule comprising a mammalian tunicamycin (Tn)-resistance gene operably linked to a gene encoding the protein substrate in need of glycosylation is provided; the cell is cultured in the presence of a first concentration of Tn; a cell population expressing at least one copy of the Tn-resistance gene is isolated; the cell population is cultured in the presence of increasing concentrations of Tn; and the N-glycan protein substrate is isolated from the cell culture. The N-glycan content of the protein substrate may be evaluated for the presence of monosaccharides and oligosaccharides by any method known in the art.

[00125] Detailed structural analysis of glycan-linked proteins may be correlated to functional features of the protein. Such analysis characterizing protein glycosylation typically involves several steps: i) an enzymatic or chemical release of the attached glycans; ii) derivatization of the released glycans via reductive amination with aromatic or aliphatic amines or permethylation; iii) analysis of the glycans. Many variations of analyzing glycosylation patterns in known to the skilled person. Glycoproteins may carry several types of glycoforms occupying various sites in specific quantities, and therefore their complexity may make it difficult

to reproduce in certain production methods. Consistency of type and quantity of glycoform is measurable and represents a desirable outcome for therapeutic protein production.

Host Cells and Transfection

[00126] The mammalian host cells used in the methods of the invention are eukaryotic host cells, usually mammalian cells, including, *e.g*. CHO cells and mouse cells. In one embodiment, the invention provides a cell comprising a nucleic acid sequence that encodes a Tn resistance marker protein derived from *Cricetulus griseus* (Chinese hamster) (as set forth in SEQ ID NO:3), or a homolog or variant thereof. In some embodiments, the cell comprises multiple gene copies of the Tn resistance marker gene. In other embodiments, the invention provides a nucleic acid sequence that encodes a Tn resistance marker protein derived from human (SEQ ID NO:4), Rhesus monkey (SEQ ID NO:5), chimpanzee (SEQ ID NO:6), dog (SEQ ID NO:7), guinea pig (SEQ ID NO:8), rat (SEQ ID NO:9) or mouse (SEQ ID NO:10).

[00127] The invention includes a mammalian host cell transfected with an expression vector of the invention. Transfected host cells include cells that have been transfected with expression vectors that comprise a sequence encoding a protein or polypeptide of interest. Expressed proteins will typically be secreted into the culture medium, depending on the nucleic acid sequence selected, but may be retained in the cell or deposited in the cell membrane. Various mammalian cell culture systems can be employed to express recombinant proteins. Examples of suitable mammalian host cell lines include the COS-7 lines of monkey kidney cells, described by Gluzman (1981) Cell 23:175, and other cell lines capable of expressing an appropriate vector including, for example, CV-1/EBNA (ATCC CRL 10478), L cells, C127, 3T3, CHO, HeLa and BHK cell lines. Other cell lines developed for specific selection or amplification schemes will also be useful with the methods and compositions provided herein. In one embodiment of the invention, the cell is a CHO cell line designated K1 (i.e. a CHO K1 cell). In order to achieve the goal of high volume production of recombinant proteins, the host cell line should be pre-adapted to bioreactor medium in the appropriate case.

[00128] Several transfection protocols are known in the art, and are reviewed in Kaufman (1988) Meth. Enzymology 185:537. The transfection protocol chosen will depend on the host cell type and the nature of the GOI, and can be chosen based upon routine experimentation. The basic requirements of any such protocol are first to introduce DNA encoding the protein of interest into a suitable host cell, and then to identify and isolate host cells which have incorporated the heterologous DNA in a relatively stable, expressible manner.

[00129] Certain reagents useful for introducing heterologous DNA into a mammalian cell include Lipofectin™ Reagent and Lipofectamine™Reagent (Gibco BRL, Gaithersburg, Md.). Both of these reagents are commercially available reagents used to form lipid-nucleic acid complexes (or liposomes) which, when applied to cultured cells, facilitate uptake of the nucleic acid into the cells.

[00130] The transfection protocol chosen and the elements selected for use therein will

depend on the type of host cell used. Those of skill in the art are aware of numerous different protocols and host cells, and can select an appropriate system for expression of a desired protein, based on the requirements of the cell culture system used. In a further aspect, the invention relates to an expression vector encoding a polypeptide, including but not limited to, an antibody, bispecific antibody, chimeric antibody, ScFv, antigen-binding protein, or Fc fusion protein. Such expression vectors may be used for recombinant production of polypeptides using the methods and compositions of the invention.

[00131] Other features of the invention will become apparent in the course of the following descriptions of exemplary embodiments which are given for illustration of the invention and are not intended to be limiting thereof.

EXAMPLES

[00132] The following examples are put forth so as to provide those of ordinary skill in the art how to make and use the methods and compositions described herein, and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.,* amount, temperature, *etc*.) but some experimental error and deviation should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Example 1. Selection efficiency of transfectant cells expressing GPT.

[00133] Modified CHO K1 cells were transfected with a plasmid vector containing CHO-GPT (SEQ ID NO: 2), human GPT (SEQ ID NO:12) or a plasmid vector containing a hygromycin phosphotransferase (Hpt, Hygromycin resistant gene); *e.g*. the selectable marker gene (CHO-GPT or hpt) was transcriptionally linked to a downstream eGFP gene via an IRES sequence, in their respective vectors. For example, each plasmid was constructed to contain the following gene sequences, in 5’ to 3’ direction: a Lox site, a SV40 late promoter, either CHO-GPT (or Hpt), IRES, enhanced green fluorescent protein (eGFP), and a second Lox site. Purified recombinant plasmids were transfected together with a plasmid that expresses Cre recombinase, into a modified CHO host cell line containing: from 5' to 3', a lox site, YFP, and a second lox site, at a transcriptionally active locus. Consequently, the host CHO cell can be isolated by flow cytometry as a green-positive or a yellow-negative cell. When the recombinant plasmid expressing eGFP (transcriptionally regulated by the GPT or hpt genes) was transfected together with a plasmid expressing the Cre recombinase, recombination mediated by the Cre recombinase results in the site-specific integration of the GPT/eGFP cassette at the chromosomal locus containing the lox sites and replacement of the YFP gene occurs (*i.e*. a green-positive cell). Should the eGFP integrate randomly, both green-positive and yellow positive cells will result.

[00134] Cell populations were either incubated with 0, 1µg/ml, 2.5µg/ml or 5µg/ml tunicamycin

(Tn) or 400µg hygromycin (Hyg), as outlined in Table 2. Observed recombinant populations (ORPs) were measured by fluorescent-activated cell sorting (FACS) analysis. Cells were sorted to quantitate each population of cells, and selection efficiency was calculated for cells expressing only GFP and not YFP (Figs. 4 or 5).

[00135] Selection efficiency (percent population of surviving cells expressing GFP) was compared between cell pools resistant to either Tn or Hyg (Table 2).

TABLE 2: Selection Efficiency

Hpt or GPT	Cre	Selection agent (ug/ml Hyg or Tn)	Selection efficiency %(Total GFP+)
Hpt	+	-	1.35
Hpt	+	+ (400 Hyg)	98.8
choGPT	+	-	0.89
choGPT	+	+ (1Tn)	86.9
choGPT	+	+ (2.5 Tn)	96.1
hGPT	+	-	2.6
hGPT	+	+ (1Tn)	97
hGPT	+	+ (2.5 Tn)	96.7

[00136] It was observed that tunicamycin selection is as efficient as hygromycin selection. . Both CHO-GPT and human GPT were efficient at selection of integrants in the presence of 1 ug/ml or 2.5 ug/ml Tunicamycin.

Example 2. Amplification of the gene product.

[00137] Incremental selection was done by applying increasing concentrations of tunicamycin to the GPT expression system. CHO K1 cells were transfected with a plasmid vector containing the CHO-GPT gene (SEQ ID NO: 2) as above. The plasmid contains in 5' to 3' direction, a first Lox site, a SV40 late promoter, the CHO-GPT gene, an IRES, eGFP, and a second Lox site. The CRE-lox sites direct integration of the gene of interest into the genome resulting in a stable transfectant pool of cells with at least one GPT insertion per cell. (More integrants may occur due to random integration, as seen above). CHO cells were initially cultured in the presence of 1

μg/ml tunicamycin (Tn). Transfectants were then selected from the stable pool (named Cell Pool 2) and subsequently expanded in the presence of 1 μg/ml, 2.5 μg/ml or 5 μg/ml Tn. Selection rounds were conducted to identify cell populations capable of enhanced expression (multiple copies) of eGFP. The random integration events increased greatly, in the presence of 2.5 μg/ml or 5 μg/ml Tn.

[00138] Copy number of gene product, either CHO GPT, eGFP or mGapdh (normalized control), was measured using standard qPCR methods. Copy number of eGFP in the cells from the 1 μg/ml Tn-resistant pool incubated further with 2.5 ug/ml Tn was at least 2 times the copy number of eGFP from the 1 μg/ml Tn-resistant pool incubated further with 1 μg/ml Tn. The gene copy number increased further when a 1 μg/ml Tn-treated pool was incubated further with 5 μg/ml Tn. The increase in gene copy number for eGFP correlates with the increased gene copy of CHO-GPT. (See Figures 6A and 6B.)

[00139] To determine whether increase in gene copy number translates to increased protein expression, the mean fluorescent intensity (MFI) was measured by FACS for the same cell pools expressing GPT and eGFP that were treated with multiple rounds of Tn selection, namely 1, 2.5 or 5 μg Tn (see *e.g.* samples 7, 8, and 9 in Figure 6B). The comparison of eGFP expression for these cell pools is represented in Table 3.

[00140] The cell pool expressing GPT that was subjected to a second round of selection with 5μg Tn resulted in just greater than 2.5 times the productive output compared to 1μg Tn treatment, and 1.5 times the productive output compared to 2.5μg Tn treatment, with respect to eGFP production (Table 3).

TABLE 3: eGFP protein production

GPT 1ug pool + Second Tn (μg) Treatment	**MFI**
1 μg	1098
2.5 μg	1867
5 μg	2854

[00141] Without being bound by any one theory, incremental increases in the concentration of Tn amplified the selective pressure to the cells in a controlled manner, thus increasing the productive output.

[00142] Tn-resistant expression vectors were also employed in further experiments, described below, to test the effects of Tn selection on glycosylation patterns.

Example 3. Expression and glycosylation profile of an exemplary dimeric protein.

[00143] CHO cells expressing a “Trap” protein (Fc fusion protein-1, hereinafter referred to as FcFP1) were transfected with a GPT-containing expression vector. The plasmid has, in 5’ to 3’ direction, a Lox site, a SV40 late promoter, a Tn-resistant gene (CHO-GPT), an IRES eGFP,

SV40 polyA and a second Lox site. 1 µg/mL Tn or 5 µg/mL Tn was used for selection of the GPT selectable marker. The selected pools cells were expanded in suspension cultures in serum-free production medium. GPT transfection was confirmed by expression of eGFP by FACS analysis. Pellets collected from selected pools were sent for copy number analysis for GPT expression and a 12 day productivity assay was set up to determine the expression level of FcFP1 in pools selected with different concentrations of Tunicamycin.

[00144] FcFP1 was selected for its complex glycosylation pattern, having an abundance of glycosylation sites. To determine glycosylation profiles, cells expressing FcF1 protein were expanded in cell culture under standard protocol (no Tn) or conditions of Tn treatment as represented in Table 4, then protein was isolated and purified.

TABLE 4: FcFP1 protein production

Protein	Lot#	Treatment
FcFP1 Trap	110728	None
FcFP1 Trap	110728-1	1 µg/ml Tn
FcFP1 Trap	110728-2	5 µg/ml Tn

[00145] Detailed glycan analysis was performed using chromatography based on well-known methods for HPLC and fluorescent anthranilic acid (AA) tags (Anumula, and Dhume, *Glycobiology*, 8(7):685-694, 1998), for each lot of glycoprotein to determine whether Tn had a negative impact on glycosylation profiles. The production lots were also compared to a reference standard which represents a therapeutically acceptable batch of protein. Representative glycan analysis is shown in Figures 7A-7D. Each lot, compared to the reference lot, consistently produces the same number of peaks, relative shape and relative intensity. An overlap of each chromatogram (Figure 8) indicates that no unique or unusual peaks are uncovered.

[00146] Oligosaccharide profiling was done by well-known HPLC methods against the reference standard lot for the FcFP1 protein. Levels of sialylation were measured for the FcFP1 trap protein lots and the Z-number was calculated for each lot (3 replicates). Z-number represents the measure of variation between lots. The Z-number takes into account the relative number of peaks, as well as the relative shape and intensity of each peak. For example, the area of each 0SA, 1SA, 2SA, 3SA and 4SA peak in Figures 7A-7D is quantitated as in Table 5.

TABLE 5: Quantitative Oligosaccharide Analysis

Protein Lot	**Replicate**	OSA	1SA	2SA	3SA	4SA	OS PROFILE (Z-number)
Reference B100002M410	1	6506.43	13388.34	11268.60	5176.21	1728.15	1.53
	2	5869.80	11932.32	10159.21	4196.10	1550.09	1.51
	3	6870.18	14536.84	12090.21	5200.58	1707.74	1.51

	Avg	6415.47	13285.83	11172.65	4857.63	1661.99	1.52±0.01
FcFP1 Trap 110728	1	6159.09	9394.92	7368.03	3074.66	675.48	1.34
	2	7530.49	12117.03	9589.08	2951.63	810.09	1.36
	3	5508.95	8580.56	6902.59	3794.81	630.79	1.34
	Avg	6399.51	10030.84	7953.23	3074.66	705.45	1.35±0.01
FcFP1 trap 110728-1	1	5330.22	8149.81	6539.33	2490.06	641.37	1.35
	2	5034.39	9009.42	7059.61	2698.05	812.21	1.40
	3	6222.44	10235.08	8428.04	3276.75	848.83	1.39
	Avg	5529.02	9131.44	7342.33	2821.62	767.47	1.38±0.03
FcFP1 trap 110728-2	1	6300.77	10001.93	8109.12	3000.96	790.99	1.36
	2	5999.09	9952.47	7968.58	2885.50	717.70	1.36
	3	4322.29	6176.33	5187.48	1742.26	458.52	1.32
	Avg	5540.72	8710.24	7088.39	2542.91	655.74	1.35±.0.02

OS = oligosaccharide; 0SA = zero sialic acid residues; 1SA = one sialic acid residue; 2SA = two sialic acid residues; 3SA = three sialic acid residues; 4SA = four sialic acid residues

$$Z\ \text{number} = \frac{(\text{Area1SA}+2*\text{Area2SA}+3*\text{Area3SA}+4*\text{Area4SA})}{(\text{Area0SA}+\text{Area1SA}+\text{Area2SA}+\text{Area3SA}+\text{Area4SA})}$$

[00147] The Z-number calculated for each lot is within an acceptable range compared to the reference lot, therefore each protein lot is understood to achieve the same material as the therapeutic molecule. Since the presence of Tn is known to have a negative effect on glycosylation of *N*-linked glycoproteins, it was unexpected that protein production would be reliable and consistent, as well as productive, given the conditions of increased selection pressure by Tn.

[00148] The present invention may be embodied in other specific embodiments without departing from the spirit or essence thereof.

Claims

1. An isolated mammalian cell comprising an exogenously introduced nucleic acid that comprises

a mammalian tunicamycin (Tn)-resistance gene encoding a protein having at least 93% identity to the amino acid sequence of SEQ ID NO: 3, and

a first gene of interest (GOI) encoding an antibody light chain or antigen-binding fragment thereof, an antibody heavy chain or antigen-binding fragment thereof, or an Fc-fusion protein,

wherein the Tn-resistance gene is operably linked to the first GOI.

2. The isolated cell of claim 1, wherein the Tn-resistance gene comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, and SEQ ID NO: 17.

3. The isolated cell of claim 1, wherein the first GOI is operably linked to at least one regulatory element selected from the group consisting of a promoter, a ribosome-binding site, and an enhancer.

4. The isolated cell of claim 1, further comprising a second exogenously-added gene of interest (GOI).

5. The isolated cell of claim 4, wherein the second GOI encodes a glycoprotein.

6. The isolated cell of claim 5, wherein the glycoprotein is selected from an antibody light chain or antigen-binding fragment thereof, an antibody heavy chain or antigen-binding fragment thereof, an Fc-fusion protein, a ligand, and a receptor or ligand-binding fragment thereof.

7. The isolated cell of claim 1, wherein the cell is selected from the group consisting of CHO, COS-7, HEK293, tumor cell, lymphocyte, retinal cell, and stem cell.

8. The isolated cell of claim 1, wherein the cell is a CHO cell.

9. A vector comprising a nucleic acid, wherein the nucleic acid comprises

a mammalian tunicamycin (Tn)-resistance gene encoding a protein having at least 93% identity to the amino acid sequence of SEQ ID NO: 3, and

a first gene of interest (GOI) encoding an antibody light chain or antigen-binding fragment thereof, an antibody heavy chain or antigen-binding fragment thereof, or an Fc-fusion protein,

wherein the Tn-resistance gene is operably linked to the first GOI.

10. The vector of claim 9, wherein the Tn-resistance gene comprises the nucleic acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16 and SEQ ID NO: 17.

11. The vector of claim 9, wherein the first GOI is operably linked to at least one regulatory element selected from the group consisting of a promoter, a ribosome-binding site, and an enhancer.

12. The vector of claim 9, further comprising a second gene of interest (GOI).

13. The vector of claim 12, wherein the second GOI encodes a glycoprotein.

14. The vector of claim 13, wherein said glycoprotein is selected from an antibody light chain or antigen-binding fragment thereof, an antibody heavy chain or antigen-binding fragment thereof, an Fc-fusion protein, a ligand, and a receptor or ligand-binding fragment thereof.

15. A method of employing tunicamycin (Tn) as a selection marker in mammalian cell culture, comprising

(a) providing a mammalian host cell population,

(b) introducing a vector according to claim 9 into the cell population of step (a) by transfection;

(c) culturing the cell population of step (b) in the presence of Tn, thereby obtaining a cell transfectant comprising said nucleic acid of said vector, wherein said first GOI of said vector encodes a first protein of interest (POI).

16. The method of claim 15, wherein the Tn is at a concentration of at least 1 μg/mL.

17. The method of claim 15, wherein the Tn is at a concentration of 1 μg/mL, 2.5 μg/mL, or 5

μg/mL.

18. The method of claim 15, wherein the cell transfectant is obtained by culturing the cell population of step (b) in the presence of sequentially increasing concentrations of Tn.

19. The method of claim 18, wherein said culturing comprising culturing at a first Tn concentration of 1 μg/mL, followed by culturing at a second Tn concentration of 2.5 μg/mL or 5 μg/mL.

20. The method of claim 15, further comprising culturing the obtained cell transfectant, expressing said first POI from said first GOI in the cell transfectant, and isolating said first POI from the cultured cell transfectant.

21. The method of claim 18, further comprising culturing the obtained cell transfectant, expressing said first POI from said first GOI in the cell transfectant, and isolating said first POI expressed from the cultured cell transfectant.

22. The method of claim 15, wherein the mammalian host cell is selected from the group consisting of CHO, COS-7, HEK293, tumor cell, lymphocyte, retinal cell, and stem cell.

23. The method of claim 15, wherein the mammalian host cell is a CHO cell.

24. The method of claim 15, wherein the vector comprises a second GOI which encodes a protein selected from the group consisting of an antibody light chain or antigen-binding fragment thereof, an antibody heavy chain or antigen-binding fragment thereof, an Fc-fusion protein, a ligand, and a receptor or ligand-binding fragment thereof.

25. A method of producing a recombinant protein of interest (POI), wherein the method comprises:

a. providing a mammalian host cell comprising an exogenous nucleic acid, wherein the nucleic acid comprises (i) a mammalian tunicamycin (Tn)-resistance gene encoding a protein having at least 93% identity to the amino acid sequence of SEQ ID NO: 3, (ii) a first a gene of interest (GOI), and (iii) at least one regulatory element, wherein the Tn-resistance gene is operably linked to the first GOI and said at least one regulatory element, and wherein said first GOI encodes a first

POI, and wherein the first POI is selected from the group consisting of an antibody light chain or antigen-binding fragment thereof, an antibody heavy chain or antigen-binding fragment thereof, and an Fc-fusion protein; and

b. culturing the cell in the presence of Tn to express said first POI.

26. The method of claim 25, wherein the Tn is at a concentration of at least 1 µg/mL.

27. The method of claim 25, wherein the Tn is at a concentration of 1 µg/mL, 2.5 µg/mL, or 5 µg/mL.

28. The method of claim 25, wherein said culturing comprising culturing in the presence of sequentially increasing concentrations of Tn.

29. The method of claim 28, wherein said culturing comprising culturing at a first Tn concentration of 1 µg/mL, followed by culturing at a second concentration of 2.5 µg/mL, or 5 µg/mL.

30. The method of claim 25, wherein the mammalian host cell is selected from the group consisting of CHO, COS-7, HEK293, tumor cell, lymphocyte, retinal cell, and stem cell.

31. The method of claim 25, wherein the mammalian host cell is a CHO cell.

32. The method of claim 25, wherein the exogenous nucleic acid further comprises a second GOI which encodes a protein selected from the group consisting of an antibody light chain or antigen-binding fragment thereof, an antibody heavy chain or antigen-binding fragment thereof, an Fc-fusion protein, a ligand, and a receptor or ligand-binding fragment thereof.

33. A method of employing tunicamycin (Tn) as a selection marker in mammalian cell culture, comprising

(a) providing a mammalian host cell population,

(b) introducing into the cell population of step (a) a nucleic acid by transfection, wherein the nucleic acid comprises (i) a mammalian tunicamycin (Tn)-resistance gene encoding a protein having at least 93% identity to the amino acid sequence of SEQ ID NO: 3, and (ii) a first a gene of interest (GOI) encoding a first protein of interest (POI), wherein the Tn-resistance gene is

operably linked to the first GOI;

(c) culturing the cell population of step (b) in the presence of sequentially increasing concentrations of Tn, thereby selecting a cell transfectant that comprises said nucleic acid; and

(d) culturing the selected cell transfectant, expressing said first POI from said first GOI in the selected cell transfectant, and isolating said first POI from the cell culture.

34. The method of claim 33, wherein the culturing in step (c) comprises culturing at a first Tn concentration of 1 μg/mL, followed by culturing at a second Tn concentration of 2.5 μg/mL or 5 μg/mL.

35. The method of claim 33, wherein the Tn-resistance gene comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, and SEQ ID NO: 17.

36. The method of claim 33, wherein the mammalian host cell is selected from the group consisting of CHO, COS-7, HEK293, tumor cell, lymphocyte, retinal cell, and stem cell.

37. The method of claim 33, wherein the mammalian host cell is a CHO cell.

38. A method of producing a recombinant protein of interest (POI), wherein the method comprises:

(a) providing a mammalian host cell comprising an exogenous nucleic acid, wherein the exogenous nucleic acid comprises (i) a mammalian tunicamycin (Tn)-resistance gene encoding a protein having at least 93% identity to the amino acid sequence of SEQ ID NO: 3, and (ii) a first a gene of interest (GOI) encoding a first POI, wherein the Tn-resistance gene is operably linked to the first GOI;

(b) culturing the cell in the presence of Tn to express said first POI; and

(c) isolating said first POI from the cell culture.

39. The method of claim 38, wherein the Tn is at a concentration of 1 μg/mL, 2.5 μg/mL, or 5 μg/mL.

40. The method of claim 38, wherein said culturing comprising culturing in the presence of sequentially increasing concentrations of Tn.

41. The method of claim 40, wherein said culturing comprising culturing at a first Tn concentration of 1 µg/mL, followed by culturing at a second concentration of 2.5 µg/mL or 5 µg/mL.

42. The method of claim 38, wherein the Tn-resistance gene comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, and SEQ ID NO: 17.

43. The method of claim 38, wherein the mammalian host cell is selected from the group consisting of CHO, COS-7, HEK293, tumor cell, lymphocyte, retinal cell, and stem cell.

44. The method of claim 38, wherein the mammalian host cell is a CHO cell.