MAMMALIAN EXPRESSION VECTORS

FIELD OF THE DISCLOSURE

The present disclosure relates to mammalian expression vectors and their use. The invention also provides a method of producing a polypeptide.

REFERENCE TO A SEQUENCE LISTING

This application includes a sequence listing.

BACKGROUND

The following introduction into the background is merely provided to aid the reader in understanding the invention and is not admitted to describe or constitute prior art to the present invention.

Mammalian cells have become the workhorse in bioprocessing of proteins for therapeutic or diagnostic use such as monoclonal antibodies. Today in upstream processing mammalian cell expression systems are the dominant production tool. Desired glycosylation patterns in combination with the development of serum-free and protein-free culture media have paved the way for this expression platform, which is characterized by a suitable cell line and an expression vector. The most commonly used cell line used worldwide is CHO. Production at large scale, combined with an improved understanding of mammalian cell nutritional requirements, have led to substantial optimizations in terms of increased efficiency and yield.

Doses of a therapeutic protein in therapeutic use are typically in the range of a few micrograms to milligrams of protein, so that further increase in efficiency and yield are desirable. Furthermore, manufacturing costs in part depend on the expression levels achieved. A critical component of a mammalian expression platform is the expression vector employed.

SUMMARY OF THE DISCLOSURE

The present disclosure can be taken to generally relate to the production of a polypeptide, such as a mammalian polypeptide. Provided is a vector suitable for expressing a respective polypeptide.

In a first aspect there is provided an expression vector. The expression vector is a mammalian expression vector. The expression vector includes a selection cassette with a eukaryotic selection marker, a selection cassette with a bacterial selection marker, an expression cassette for a target polypeptide and a bacterial origin of

replication. The selection cassette with a eukaryotic selection marker includes a nucleotide sequence that encodes a glutamine synthetase as the eukaryotic selection marker. The nucleotide sequence encoding glutamine synthetase is operably linked to a 3-phosphoglycerate kinase (PGK) promoter and a polyadenylation (pA) signal. The selection cassette with a bacterial selection marker includes a nucleotide sequence that encodes an enzyme which confers resistance against an antibiotic to a bacterial host as the bacterial selection marker, and that is operably linked to a suitable promoter. The expression cassette for a target polypeptide includes an insertion site for a nucleotide sequence that encodes the target polypeptide. The insertion site is operably linked to a cytomegalovirus (CMV) promoter and a polyadenylation (pA) signal.

According to some embodiments the expression vector is a mammalian expression vector. According to some embodiments the expression vector is a vector for a murine or a hamster cell line, such as a CHO cell line or an NS0 (non secreting) myeloma cell line. According to some embodiments the expression vector is an expression vector for CHO cells.

The CMV promoter is generally the human cytomegalovirus immediate early. It is for example found in the pcDNA3.1 or the pCMV vectors.

As noted above, the expression vector includes a selection cassette that includes a nucleotide sequence, which encodes an enzyme conferring resistance against an antibiotic to a bacterial host. The respective antibiotic is in some embodiments Ampicillin. A suitable enzyme that may be encoded as the bacterial selection marker conferring resistance against Ampicillin to a bacterial host is beta-lactamase.

According to some embodiments of the expression vector the glutamine synthetase encoded by the sequence included in the expression vector has a sequence that is 98 % or more identical to SEQ ID NO: 5 and is capable of catalyzing the ATP-dependent conversion of glutamate and ammonia to glutamine.

According to some embodiments the selection cassette with a eukaryotic selection marker includes a nucleotide sequence that encodes a mammalian glutamine synthetase as the eukaryotic selection marker. In some embodiments the glutamine synthetase has a nucleic acid sequence of at least 90 % identity to the sequence of SEQ ID NO: 3.

In some embodiments the glutamine synthetase included as a eukaryotic selection marker is a CHO glutamine synthetase. In some embodiments the glutamine synthetase has a nucleic acid sequence of at least 96 % identity to the sequence of SEQ ID NO: 3.

In some embodiments the expression vector further includes an expression augmenting sequence element (EASE). In some embodiments the expression vector further includes a selection cassette that includes a nucleotide sequence, which encodes an enzyme that confers resistance against the antibiotic puromycin as a

eukaryotic selection marker, operably linked to a 3-phosphoglycerate kinase (PGK) promoter and a polyadenylation (pA) signal. The enzyme that confers resistance against the antibiotic puromycin is in some embodiments puromycin-N-acetyltransferase (pac).

According to some embodiments of the expression vector the bacterial origin of replication is the pUC origin of replication.

According to some embodiments the PGK promoter has a sequence of at least 98 % identity to SEQ ID NO: 2 and is a functional promoter. According to some embodiments the glutamine synthetase is a CHO glutamine synthetase that has the sequence of SEQ ID NO: 3. According to some embodiments the PGK promoter includes the sequence of SEQ ID NO: 2. According to some embodiments the glutamine synthetase is a CHO glutamine synthetase that includes the sequence of SEQ ID NO: 3. According to some embodiments a sequence that includes the sequence of the respective CHO glutamine synthetase and of the 3-phosphoglycerate kinase (PGK) promoter operably linked thereto has the sequence of SEQ ID NO: 1. According to some embodiments the CMV promoter has the sequence of SEQ ID NO: 4.

According to some embodiments the pA signal is a simian virus 40 pA signal. The simian virus 40 pA signal may in some embodiments be the early pA and late pA signal.

According to some embodiments the bacterial selection marker conferring resistance against Ampicillin to a bacterial host is a beta-lactamase having an identity of 95 % to the sequence of SEQ ID NO: 6.

In a second aspect there is provided a recombinant host cell. The host cell includes the expression vector according to the first aspect.

In some embodiments the recombinant host cell is a CHO cell. In some embodiments the recombinant host cell is a NS0 cell. In some embodiments the recombinant host cell is a COS-7 monkey kidney cell. In some embodiments the recombinant host cell is a 3T3 cell. In some embodiments the recombinant host cell is a Baby hamster kidney (BHK) cell. In some embodiments the recombinant host cell is a human embryonic kidney 293 cell.

In a third aspect there is provided a method of producing a polypeptide. The method involves culturing the recombinant host cell according to second aspect under conditions suitable for expressing a heterologous target polypeptide. The recombinant host cell according to second aspect includes an expression vector, in which a nucleotide sequence encoding the polypeptide as a heterologous target polypeptide is included at the insertion site for the nucleotide sequence encoding the target polypeptide.

In typical embodiments of the method according to the third aspect the nucleotide

sequence encoding the polypeptide as a heterologous target polypeptide is operably linked to the CMV promoter and to the pA signal.

In some embodiments the method according to the third aspect includes maintaining the host cell, e.g. a CHO cell or a NS0 cell, as a suspension in a suitable medium. In some embodiments the method according to the third aspect includes maintaining the host cell, e.g. a CHO cell or a NS0 cell, at a temperature in the range from 33 to 38 °C. The temperature may for instance be chosen as about 37 °C.

In a fourth aspect there is provided the use of an expression vector according to the first aspect for expressing a polypeptide at high levels.

In a fifth aspect there is provided the use of a host cell according to second aspect for expressing a polypeptide at high levels.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a vector map of the known vector pcDNA 3.1 (+), a parent vector for GA fragment 1. **Figure 1B** is a vector map of pMBL C-GS. **Figure 1C** is a vector map of pMBL CE-GS. **Figure 1D** is a linear display of the vector pMBL CE-GS.

Figure 2A depicts the sequence of SEQ ID NO: 1, which includes the sequence of a CHO glutamine synthetase as a eukaryotic selection marker, operably linked to a PGK promoter. **Figure 2B** depicts the sequence of SEQ ID NO: 2, which is a PGK promoter. **Figure 2C** depicts the sequence of SEQ ID NO: 3, which is the open reading frame encoding a glutamine synthetase as a eukaryotic selection marker. **Figure 2D** depicts the sequence of SEQ ID NO: 4, the sequence of a CMV promoter. **Figure 2E** depicts the sequence of SEQ ID NO: 14, the sequence of an expression augmenting sequence element.

Figure 3 is a vector map of pMBL CE-Puro.

Figure 4 is a vector map of pMBL C-Puro.

Figure 5A depicts the gel electrophoretic restriction analysis of vector pMBL C-GS. **Figure 5B** depicts the gel electrophoretic restriction analysis of vector pMBL CE-GS. **Figure 5C** depicts the gel electrophoretic restriction analysis of vector pMBL CE-Puro. **Figure 5D** depicts the gel electrophoretic restriction analysis of vector pMBL C-Puro.

Figure 6 is a table providing an overview on the monoclonal antibodies expressed using vectors according to the disclosure.

Figure 7A depicts the titers of anti-CTLA4 antibody Bmab700 expressed using the vector pMBL CE-GS (*a*) and the vector pMBL C-GS (*b*), in comparison to the expression after co-transfection of HC and LC of the same anti-CTLA4 antibody in two individual pCHO 1.0 vectors (*c*) and expression using a single pCHO 1.0 vector,

into which both HC and LC had been cloned (*d*). **Figure 7B** depicts the titers of four different anti-CD20 fusion antibodies at days 5, 7, 9, 12 and 14 expressed using pMBL CE-GS. Both HC and LC had been cloned into the vector pMBL CE-GS at the XhoI - NotI sites. Days are indicated above the bars. Transient transfections were performed in co-transfection mode. All transfections were carried out by using The ExpiCHO™ Expression System (Cat. no. A29133 from Thermo Fisher Scientific Inc., Waltham, MA USA), this is a high-yield transient expression system based on suspension-adapted Chinese Hamster Ovary (CHO) cells. ExpiCHO-S cells were thawed as per the expiCHO-S expression system manual. After a minimum of 2 passages post thaw, the cells were seeded for transfection at 3.5 x 10^6 cells/ml one day before transfection. On day 1 post seeding, the cell count was estimated and adjusted to 6.0 x 10^6 cells/ml by diluting with fresh, pre-warmed medium for each transfection. Transfection was carried out as per expiCHO-S expression system manual. DNA was transfected at 1µg/ml final concentration in culture. Expifectamine and feeds were added as per manufacturer's protocol. Max titre protocol was followed for all transfections. Transfection flasks were cultivated at 37° C incubator with a humidified atmosphere of 8% CO_2 in air on an orbital shaker. VCD and viabilities were estimated on days, 0, 2, 5, 7, 9, and 12. Cultures were harvested when viability dropped below 50% or on D-12.

Figure 8A depicts the time course of the viable cell density, and **Fig. 8B** of the cell viability in % of CHO cells expressing an anti-CD20 HC-C-TGF□RII fusion antibody, (Fmab25/1 and Fmab25/2), an anti-CD20 LC-C-TGF□RII fusion antibody (Fmab26/1 and Fmab26/2), an anti-CD20 HC-N-TGF□RII fusion antibody (Fmab27/1 and Fmab27/2), and a further anti-CD20 LC-N-TGF□RII fusion antibody (Fmab28/1 and Fmab28/2), transiently expressed using the vector pMBL CE-GS. **Fig. 8C** depicts the time course of the titer of product produced. Measurements were done on day 0, day 2, day 5, day 7, day 9, day 12 and day 14. Numbers on bars indicate the respective day of culture, *a* = day 12, *b* = day 14. **Fig. 8D** shows the harvest titer of product produced on day 14. Harvest Titer analysis was done based on HPLC method for Day-14 samples. Transient expression studies were conducted as described in paragraph [0032].

Figure 9A depicts the time course of the viable cell density, and **Fig. 9B** of the cell viability in % of CHO cells expressing an anti CD-20 wild type antibody, an anti-CD20 HC-C-TGFβRII fusion antibody (Fmab25), and an anti-CD20 TIM3 LC-C-Terminus fusion antibody (Fmab30). Growth profiles were monitored on day 0, day 2, day 5, day 7 and day 9. Numbers on bars indicate the respective day of culture. **Fig. 9C** depicts the time course of the titer of product produced. Measurements were done on day 5, day 7 and day 9. Numbers on bars indicate the respective day of culture. **Fig. 9D** shows the harvest titer of product produced on day 9. Numbers on bars indicate the titer obtained. Analysis was based on HPLC.

Figure 10A depicts the time course of the viable cell density, and **Fig. 10B** of the cell

viability in % of CHO cells expressing an anti-PDL-1 antibody. Measurements were done on day 0, day 2, day 5, day 7, day 9 and day 12. Letters on bars indicate the respective antibody chain expressed: a-c: Atezolizumab, d-f: anti-PDL-1 antibody. **Fig. 10C** shows the harvest titer of product produced on day 14. Numbers on bars indicate the titer obtained. Analysis was based on HPLC.

Figure 11A depicts the time course of the viable cell density, and **Fig. 11B** of the cell viability in % CHO cells expressing a further anti-PDL-1 antibody. Measurements were done on day 0, day 5, day 9 and day 12. Numbers on bars indicate the respective day of culture. **Fig. 11C** depicts the time course of the titer of product produced. Measurements were done on day 9 and day 12. Numbers on bars indicate the respective day of culture. **Fig. 11D** shows the harvest titer of product produced on day 12. Numbers on bars indicate the titer obtained. Analysis was based on HPLC.

Figure 12A shows the viability in six flasks in % and **Fig. 12B** the viable cell density during the selection phase in the absence of glutamine. Flasks 2A, 2B, 3A and 3B contained CHO cells of different CHO-S GS knock out hosts, stably transfected with pMBL CE-GS by means of electroporation, and Flasks 5 and 6 represent mock cell controls containing CHO cells electroporated with no DNA for C12-1 and C#26-2 respectively. Flasks 2A and 2B and flasks 3A and 3B were replicate flasks of the same cell lines, respectively. pMBL CE-GS encoded etanacerpt, a fusion protein of the extracellular ligand-binding domain of the p75 TNF receptor and the human Fc portion of IgG1. Each day 0 was the day of seeding. Days 7/6, 6, 3 and 4, respectively were days in selection.

Figure 13A depicts the time course of the viable cell density, and **Fig. 13B** of the cell viability in % of CHO cells of the same flasks as in Fig 12A and B, stably expressing etanacerpt. Glucose was used as feed under simple fed batch culture, data were collected at passage No 5. **Fig. 13C** depicts the time course of the titer of product produced. Measurements were done on day 5, day 7, day 10 and day 14. Numbers on bars indicate the respective day of culture. **Fig. 13D** shows the harvest titer of product produced on day 14. Titer were estimated by using HPLC.

Figure 14A depicts the time course of the viable cell density, and **Fig. 14B** of the cell viability in % of CHO cells of the same pools as in Fig 12A, B and Fig. 13, stably expressing etanacerpt. CB4 (Cell Boost-4 from Hyclone, GE Healthcare Life Sciences) and EFC (Efficient feed C from Thermo Fisher Scientific Inc., Waltham, MA USA) were used as feeds under fed batch culture, data were collected at passage No 7. **Fig. 14C** depicts the time course of the titer of product produced. Measurements were done on day 7 and day 12. Numbers on bars indicate the respective day of culture. **Fig. 14D** shows the harvest titer of product produced on day 12.

Figure 15 depicts a Western blot analysis of harvest samples for which data are provided in Fig. 13 and Fig. 14. Non-purified harvest supernatant from fed batch (cf. Fig. 14) and simple fed batch (cf. Fig. 13) was analysed.

DETAILED DESCRIPTION OF THE INVENTION

In order that the explanations on the nucleic acid molecules, vectors, host cells, methods and uses disclosed herein may be more readily understood, certain terms are first defined.

Definitions

Unless otherwise stated, the following terms used in this document, including the description and claims, have the definitions given below.

The word "about" as used herein refers to a value being within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, "about" can mean within 1 or more than 1 standard deviation, per the practice in the art. The term "about" is also used to indicate that the amount or value in question may be the value designated or some other value that is approximately the same. The phrase is intended to convey that similar values promote equivalent results or effects according to the invention. In this context "about" may refer to a range above and/or below of up to 10%. The word "about" refers in some embodiments to a range above and below a certain value that is up to 5%, such as up to up to 2%, up to 1%, or up to 0.5 % above or below that value. In one embodiment "about" refers to a range up to 0.1 % above and below a given value.

The term "essentially consists of" is understood to allow the presence of additional components in a sample or a composition that do not affect the properties of the sample or a composition. As an illustrative example, a pharmaceutical composition may include excipients if it essentially consists of an active ingredient.

The terms "expressing" and "expression" in reference to a polypeptide are intended to be understood in the ordinary meaning as used in the art. A polypeptide is expressed by a cell via transcription of a nucleic acid into mRNA, followed by translation into a polypeptide, which is folded and possibly further processed. With regard to the respective biological process itself, the terms "expression", "gene expression" or "expressing" refer to the entirety of regulatory pathways converting the information encoded in the nucleic acid sequence of a gene first into messenger RNA (mRNA) and then to a protein. Accordingly, the expression of a gene includes its transcription into a primary hnRNA, the processing of this hnRNA into a mature RNA and the translation of the mRNA sequence into the corresponding amino acid sequence of the polypeptide. In this context, it is also noted that the term "gene product" refers not only to a polypeptide, including e.g. a final polypeptide (including a splice variant thereof) encoded by that gene and a respective precursor polypeptide where applicable, but also to the respective mRNA, which may be regarded as the "first gene product" during the course of gene expression.

The terms "expression vector" or "expression construct" refer to a nucleic acid vehicle

such as a plasmid, by means of which a desired target polypeptide can be expressed in a host cell using the transcription and translation machinery of the host cell. The nucleic acid molecule can be introduced into the respective host cell and includes one or more regulatory sequences operably linked to nucleic acid sequence that encodes the target polypeptide.

By "fragment" in reference to a polypeptide such as an immunoglobulin or a proteinaceous binding molecule is meant any amino acid sequence present in a corresponding polypeptide, as long as it is shorter than the full length sequence and as long as it is capable of performing the function of interest of the polypeptide – in the case of an immunoglobulin specifically binding to the desired target, e.g. antigen such as PDL-1. The term "immunoglobulin fragment" refers to a portion of an immunoglobulin, often the hypervariable region and portions of the surrounding heavy and light chains that displays specific binding affinity for a particular molecule. A hypervariable region is a portion of an immunoglobulin that physically binds to the polypeptide target.

The term "nucleic acid molecule" as used herein refers to any nucleic acid in any possible configuration, such as single stranded, double stranded or a combination thereof. Examples of nucleic acids include for instance DNA molecules, RNA molecules, analogues of the DNA or RNA generated using nucleotide analogues or using nucleic acid chemistry, locked nucleic acid molecules (LNA), protein nucleic acids molecules (PNA), alkylphosphonate and alkylphosphotriester nucleic acid molecules and tecto-RNA molecules (e.g. Liu, B., et al., J. Am. Chem. Soc. (2004) 126, 4076-4077). LNA has a modified RNA backbone with a methylene bridge between C4' and O2', providing the respective molecule with a higher duplex stability and nuclease resistance. Alkylphosphonate and alkylphosphotriester nucleic acid molecules can be viewed as a DNA or an RNA molecule, in which phosphate groups of the nucleic acid backbone are neutralized by exchanging the P-OH groups of the phosphate groups in the nucleic acid backbone to an alkyl and to an alkoxy group, respectively. DNA or RNA may be of genomic or synthetic origin and may be single or double stranded. Such nucleic acid can be e.g. mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, a copolymer of DNA and RNA, oligonucleotides, etc. A respective nucleic acid may furthermore contain non-natural nucleotide analogues and/or be linked to an affinity tag or a label.

Many nucleotide analogues are known and can be used in nucleic acids used in the methods of the invention. A nucleotide analogue is a nucleotide containing a modification at for instance the base, sugar, or phosphate moieties. As an illustrative example, a substitution of 2'-OH residues of siRNA with 2'F, 2'O-Me or 2'H residues is known to improve the in vivo stability of the respective RNA. Modifications at the base moiety may be a natural or a synthetic modification of A, C, G, and T/U, a different purine or pyrimidine base, such as uracil-5-yl, hypoxanthin-9-yl, and 2-aminoadenin-9-yl, as well as a non-purine or a non-pyrimidine nucleotide base. Other

nucleotide analogues serve as universal bases. Examples of universal bases include 3-nitropyrrole and 5-nitroindole. Universal bases are able to form a base pair with any other base. Base modifications often can be combined with for example a sugar modification, such as for instance 2'-O-methoxyethyl, e.g. to achieve unique properties such as increased duplex stability.

The terms "polypeptide" and "protein" refer to a polymer of amino acid residues and are not limited to a certain minimum length of the product. Where both terms are used concurrently, this twofold naming accounts for the use of both terms side by side in the art.

The terms "polyadenylation site", "poly A site" or "poly A sequence" as used herein refers to a nucleic acid sequence such as a DNA sequence that allows directing both the termination and polyadenylation of a nascent RNA transcript. Efficient polyadenylation of a recombinant transcript is advantageous as a transcript lacking a poly A tail is usually unstable and rapidly degraded. The poly A signal utilized in the vector disclosed herein may be of any origin and may also be "endogenous". An endogenous poly A signal is one that is found naturally at the 3' end of the coding region of a given gene in a genome. A commonly used heterologous poly A signal is the SV40 poly A signal. Examples of suitable polyadenylation sequences also include, but are not limited to the bovine growth hormone (bGH) polyadenylation signal, a beta-globin polyA site, and a herpes simplex virus thymidine kinase polyA site.

The term "purified" is understood to be a relative indication in comparison to the original environment of the cell, thereby representing an indication that the cell is relatively purer than in the natural environment. It therefore includes, but does not only refer to, an absolute value in the sense of absolute purity from other cells (such as a homogeneous cell population). Compared to the natural level, the level after purifying the cell will generally be at least 2-5 fold greater (e.g., in terms of cells/ml). Purification of at least one order of magnitude, such as about two or three orders, including for example about four or five orders of magnitude is expressly contemplated. It may be desired to obtain the cell at least essentially free of contamination, in particular free of other cells, at a functionally significant level, for example about 90%, about 95%, or 99% pure. With regard to a nucleic acid, peptide or a protein, the above applies mutatis mutandis. In this case purifying the nucleic acid, peptide or protein will for instance generally be at least 2-5 fold greater (e.g., in terms of mg/ml).

The word "recombinant" is used in this document to describe a nucleic acid molecule that, by virtue of its origin, manipulation, or both is not associated with all or a portion of the nucleic acid molecule with which it is associated in nature. Generally a recombinant nucleic acid molecule includes a sequence which does not naturally occur in the respective wild type organism or cell. Typically a recombinant nucleic acid molecule is obtained by genetic engineering, usually constructed outside of a cell.

Generally a recombinant nucleic acid molecule is substantially identical and/or substantial complementary to at least a portion of the corresponding nucleic acid molecule occurring in nature. A recombinant nucleic acid molecule may be of any origin, such as genomic, cDNA, mammalian, bacterial, viral, semisynthetic or synthetic origin. The term "recombinant" as used with respect to a protein / polypeptide means a polypeptide produced by expression of a recombinant polynucleotide.

The terms "comprising", "including," containing", "having" etc. shall be read expansively or open-ended and without limitation. Singular forms such as "a", "an" or "the" include plural references unless the context clearly indicates otherwise. Thus, for example, reference to a "vector" includes a single vector as well as a plurality of vectors, either the same - e.g. the same operon - or different. Likewise reference to "cell" includes a single cell as well as a plurality of cells. Unless otherwise indicated, the term "at least" preceding a series of elements is to be understood to refer to every element in the series. The terms "at least one" and "at least one of" include for example, one, two, three, four, or five or more elements. It is furthermore understood that slight variations above and below a stated range can be used to achieve substantially the same results as a value within the range. Also, unless indicated otherwise, the disclosure of ranges is intended as a continuous range including every value between the minimum and maximum values.

The scope and meaning of any use of a term will be apparent from the specific context in which the term is used. Certain further definitions for selected terms used throughout this document are given in the appropriate context of the detailed description, as applicable. Unless otherwise defined, all other scientific and technical terms used in the description, figures and claims have their ordinary meaning as commonly understood by one of ordinary skill in the art.

Provided are vectors, i.e. circular nucleic acid molecules, that are capable of driving high levels of heterologous protein expression in mammalian cells, in particular in CHO cells, as well as methods and uses based thereon. The vectors are replicable in bacterial cells.

Being an expression vector, a vector according to the present disclosure generally includes a promoter, a transcription terminator sequence, an origin of replication, a selectable marker, an insertion site for a nucleotide sequence and a regulatory element.

A promoter is a region of DNA that initiates transcription of a particular gene. A cytomegalovirus promoter is an example of a mammalian expression promoter, being derived from cytomegalovirus. The CMV promoter is commonly used in vectors used in genetic engineering work conducted in mammalian cells, as it is a strong promoter and drives constitutive expression of genes under its control. A promoter is operably linked to the insertion site for a nucleotide sequence encoding the target polypeptide.

A transcription terminator is a portion of nucleic acid sequence that marks the end of a gene or operon in genomic DNA for the transcription process. Such a sequence mediates transcriptional termination by providing a signal in the newly synthesized mRNA that triggers a process which releases the mRNA from the transcriptional complex.

A transcription terminator sequence may be operably linked to the insertion site for a nucleotide sequence encoding the target polypeptide.

An origin of replication, also called replication origin, is a particular sequence in a genome at which replication is initiated. DNA replication typically begins at a single origin of replication. In *E. coli*, the origin of replication is called oriC. This region is included into an expression vector capable of replicating itself in *E.coli* to form multiple copies.

A selection marker is a gene introduced into a cell, such as a bacterium that confers a trait suitable for artificial selection.

Transcription of eukaryotic genes is regulated by a variety of cis- and trans- acting regulatory elements. Two of the best characterized cis elements are promoters and enhancers. A promoter, see above, is a nucleic acid sequence immediately 5‘ to the coding sequence of a gene; it includes multiple binding sites for trans-acting transcription factors, forming the basal transcription apparatus. Enhancers generally include multiple binding sites for trans-acting transcription factors but they can be found far upstream or downstream of a coding sequence or even within an intron. A cis element that is included in some embodiments of the vector disclosed herein is the Expression Augmentation Sequence Element (EASE). The EASE element contributes to the stability of the vector which is important for stable cell line development.

An expression vector as disclosed herein includes three distinct expression cassettes. A first cassette is an expression cassette for expressing a selection marker protein that allows the selection of eukaryotic cells that include the vector. This expression cassette is herein also referred to as a selection cassette. A second cassette is an expression cassette for expressing a selection marker protein that allows the selection of bacterial cells that include the expression vector. This expression cassette is herein likewise also referred to as a selection cassette. A third cassette is an expression cassette for expressing a target polypeptide. This expression cassette generally includes a site for inserting a nucleotide sequence that encodes the target polypeptide downstream and in operable linkage to a promoter of the expression cassette.

The insertion site of the third cassette typically contains at least one restriction enzyme recognition sequence. It may include two or more restriction enzyme recognition sequences and define a multiple cloning site. In some embodiments the insertion site includes a recognition sequences for the *Not* I enzyme. In some embodiments the insertion site includes a recognition sequences for the *Xho* I enzyme. In some

embodiments the insertion site includes a recognition sequences for the *BamH* I enzyme. In some embodiments the insertion site includes a recognition sequences for the *Nhe* I enzyme. In some embodiments the insertion site includes a recognition sequences for the *EcoR* I enzyme. In some embodiments the insertion site includes a recognition sequences for the EcoR V enzyme. In some embodiments the insertion site includes a recognition sequences for the *Pme* I enzyme. In some embodiments the insertion site includes a recognition sequences for the *Afl* II enzyme. In some embodiments the insertion site includes a recognition sequences for the *Hind* III enzyme. In some embodiments the insertion site includes a recognition sequences for the *Kpn* I enzyme. In some embodiments the insertion site includes a recognition sequences for the *Xba* I enzyme. In some embodiments the insertion site includes a recognition sequences for the *BstX* I enzyme. In some embodiments the insertion site includes a recognition sequences for the *Pme* I enzyme. Cleavage of the circular vector using any one or two of the enzymes for which there is a restriction enzyme recognition sequence creates a linear vector to which a nucleotide sequence encoding the target polypeptide with appropriate ends may be attached.

The three expression cassettes may be arranged in the vector in any order relative to each other. In some embodiments the first expression cassette for expressing a selection marker protein that allows the selection of eukaryotic cells and the second expression cassette for expressing a selection marker protein that allows the selection of bacterial cells are arranged in the vector in opposite directions. In some embodiments the first expression and the second expression cassette are arranged in the vector in the same direction. In some embodiments the first expression cassette for expressing a selection marker protein that allows the selection of eukaryotic cells and the third expression cassette for expressing a target polypeptide are arranged in the vector in the same direction. In some embodiments the first and the third expression cassette are arranged in the vector in opposite directions. In some embodiments the second expression cassette for expressing a selection marker protein that allows the selection of bacterial cells and the third expression cassette for expressing a target polypeptide are arranged in the vector in opposite directions. In some embodiments the second and the third expression cassette are arranged in the vector in the same direction.

An illustrative example of an order and orientation of the three expression cassettes is shown in Figures 1B, 1C, 3 and 4.

A eukaryotic selection marker present in a vector according to the present disclosure is a nucleic acid sequence encoding glutamine synthetase. The enzyme glutamine synthetase is the enzyme responsible for the biosynthesis of glutamine from glutamate and ammonia. This enzymatic reaction provides the only pathway for the synthesis of glutamine in a mammalian cell. In the absence of glutamine in the growth medium, the GS enzyme is essential for the survival of mammalian cells in culture. Mammalian cell lines, such as a Chinese hamster ovary (CHO) cell line, express sufficient GS to

survive without exogenous glutamine. GS is the most commonly used selection marker for selecting transfectants while creating stable cell lines. For a cell line in which the endogenous GS in the host is not functional, because of an inactivating mutation or a deletion, supplementation of exogenous GS through the vector allows selection of a transfected population containing the vector of interest. The resultant recombinant cell line can be screened in a glutamine-free medium thereby reducing time, and increasing the probability of obtaining a high titer clone.

On a vector disclosed herein, the enzyme glutamine synthetase is encoded by a nucleic acid sequence that is operably linked to a PGK promoter. PGK is in some embodiments PGK-1. The PGK-1 gene encodes an enzyme of the glycolytic pathway, the housekeeping enzyme, 3-phosphoglycerate kinase. It is therefore ubiquitously expressed. This gene is located on the X chromosome in mammals. In the somatic cells of female mammals only the PGK-1 allele on the active X chromosome is transcribed, the other PGK-1 allele on the inactive X being inert. Hence, PGK-1 is always expressed except for the situation when it is silenced along with most other genes on the inactive X chromosome of female somatic cells or male germ cells. The PGK-1 promoter is thus used for gene expression at high constitutive levels, since it is active in virtually all somatic and germ cell types.

In some embodiments the PGK-1 promoter has the sequence of SEQ ID NO: 2. The PGK-1 promoter of SEQ ID NO: 2 is for example included in a vector depicted in Table 2 of WO 2014/200557. It is also included in the cloning vector PGK1p-Csy4-pA of GenBank accession number KJ796485.1, version 1 dated 16 August 2014 or in the cloning vector PBDGTV of GenBank accession number KU179219.1, version 1 dated 17 January 2017.

In some embodiments the PGK promoter has a sequence that is at least 98 % identical to the sequence of SEQ ID NO: 2. As an illustrative example, the mutation introducing vector pMtKCNQ2 DNA of GenBank accession number AB535097.1, version 1 as of 4 December 2009, has a PGK promoter sequence in positions 2984 to 3491 that lacks bases 371 to 382 of SEQ ID NO: 2. As another example, the mPGK1 promoter at positions 6435 to 6924 of the vector pROSA26-DV3 of GenBank accession number LT726831.1, version 1 of 6 February 2017, has six single base deletions relative to the sequence of SEQ ID NO: 2. In some embodiments the PGK promoter has a sequence that is at least 99 % identical to the sequence of SEQ ID NO: 2. For example the mouse phosphoglycerate kinase promoter at positions 6566 to 7056 of the plasmid vector pHM2 of GenBank accession number X76683.1, version 1 of 9 February 1994, has five deleted bases and one substitution relative to the sequence of SEQ ID NO: 2. As a further example, the complementary sequence of the PGK-1 promoter at positions 4395 to 4888 of the cloning vector pGZ-DSB-CO of GenBank accession number KY447298.1, version 1 of 22 February 2017, has a deletion of two bases relative to the sequence of SEQ ID NO: 2.

The sequence encoded as CHO glutamine synthetase is in some embodiments the sequence of SEQ ID NO: 5. The sequence is found in GenPept under accession number AJHYQ of 3 June 2002. The sequence encoded as CHO glutamine synthetase is in some embodiments a sequence of 99 % or more identity to SEQ ID NO: 5. The sequence may for instance be the sequence encoded by the sequence of SwissProt/UniProt accession number G3HG36, version 1 of the sequence of 16 November 2011, version 30 of the entry of 23 May 2018, which is the sequence encoded by the sequence of GenBank accession number RLQ66161.1, version 1 of 21 October 2018. In some embodiments the sequence encoded as CHO glutamine synthetase is a sequence of 97 % or more identity to SEQ ID NO: 5. The sequence encoded as CHO glutamine synthetase is in some embodiments a sequence of 96 % or more identity to SEQ ID NO: 5. As an illustrative example, the sequence may be the sequence of SwissProt/UniProt accession number P04773, version 4 of the sequence of 23 January 2007, version 107 of the entry of 5 December 2018. The sequence encoded as CHO glutamine synthetase includes in some embodiments the sequence of SEQ ID NO: 5. The sequence encoded as CHO glutamine synthetase includes in some embodiments a sequence of 99 % identity to SEQ ID NO: 5. The sequence encoded as CHO glutamine synthetase includes in some embodiments a sequence of 99 % or more identity to SEQ ID NO: 5. The sequence encoded as CHO glutamine synthetase includes in some embodiments a sequence of 97 % or more identity to SEQ ID NO: 5. The sequence encoded as CHO glutamine synthetase includes in some embodiments a sequence of 96 % or more identity to SEQ ID NO: 5. In one embodiment the sequence may be the sequence of SwissProt/UniProt accession number G3IH33, version 1 of the sequence of 16 November 2011, version 22 of the entry of 5 December 2018.

The sequence encoding CHO glutamine synthetase has in some embodiments the sequence of SEQ ID NO: 3. The sequence encoding CHO glutamine synthetase of SEQ ID NO: 3 is for example included, as positions 147 to 1268, in the sequence of mRNA encoding Chinese hamster glutamine synthetase of GenBank accession number X03495.1, version 1 dated 21 April 1993, the database entry being last updated on 4 February 2011. As a further example, the sequence is found in WO 2013/186371 as SEQ ID NO: 10 or as SEQ ID NO: 1 in EP 2 825 641.

The glutamine synthetase (GS) Gene Expression System (Birch J.R. and Racher A. J., Advanced Drug Delivery Reviews 2006; 58:671-685) is one of two commonly used expression vector systems in monoclonal antibody production. Another common expression system is based on dihydrofolate reductase (DHFR) genes. The GS system is particularly useful for CHO and NS0 cells, being based on the metabolic pathway of glutamate and ammonium to glutamine for the selection of recombinant cells. CHO cells already express endogenous GS. The addition of a selective GS inhibitor, such as methionine sulphoximine (MSX), to a glutamine-free culture medium selects cell clones having integrated the gene construct containing the GS gene.

The sequence encoding glutamine synthetase, e.g. a CHO glutamine synthetase has in some embodiments a sequence that is at least 97 % identical to the sequence of SEQ ID NO: 3. The sequence of CHO glutamine synthetase of NCBI accession number NM_001246770.1, version 1 of 9 October 2011, contains at positions 1 to 1116 a promoter sequence that differs in 39 substituted bases from the sequence of SEQ ID NO: 3. In some embodiments the sequence encoding CHO glutamine synthetase has a sequence that is at least 92 % identical to the sequence of SEQ ID NO: 3. As an example, the sequence of murine glutamine synthetase of GenBank accession number X16314.1, version 1 of 4 April 1995, has 91 base substitutions relative to the sequence of SEQ ID NO: 3. The sequence encoding CHO glutamine synthetase has in some embodiments a sequence of 90 % identity to SEQ ID NO: 3. As an example, the human glutamine synthetase of GenBank accession number BC051726, version 1 of 14 May 2003, has a sequence in positions 1307 to 2422 that differs in 106 substitutions from the sequence of SEQ ID NO: 3. As a further example, the porcine glutamine synthetase of GenBank accession number AK390323.1, version 1 of 11 January 2012, has a sequence in positions 252 to 1367 that differs in 116 substitutions from the sequence of SEQ ID NO: 3.

The vector of the present disclosure includes a bacterial selection marker. This marker is a nucleotide sequence that encodes an enzyme that provides resistance against an antibiotic to a bacterial host. In some embodiments the enzyme provides resistance to chloramphenicol or to kanamycin and geneticin. In some embodiments the enzyme provides resistance to ampicillin. In some embodiments the enzyme provides resistance to streptomycin and spectinomycin.

An example of an enzyme that provides resistance to a bacterial host against ampicillin is beta-lactamase. In some embodiments the enzyme beta-lactamase has the amino acid sequence of SEQ ID NO: 6, which is the sequence of beta-lactamase of inter alia *Escherichia coli* of SwissProt/UniProt accession number Q79DR3, version 1 of the sequence of 5 July 2004, version 116 of the entry of 5 December 2018. The sequence is also identical to the synthetic beta-lactamase construct of SwissProt/UniProt accession number Q285M4, version 1 of the sequence of 4 April 2006, version 36 of the entry of 10 May 2017. In some embodiments the beta-lactamase may be a protein with 99 % or more identity to the sequence of SEQ ID NO: 6. It may for example be the beta-lactamase protein encoded by the plasmid pPV with the amino acid sequence of the SwissProt/ UniProt accession number Q799Y1, version 1 of the sequence of 5 July 2004, version 42 of the entry of 10 May 2017. It may also be the class A broad-spectrum beta-lactamase TEM-1 from *Bordetella avium* with the amino acid sequence of the SwissProt/UniProt accession number A0A3A0YVF2, version 1 of the sequence of 5 December 2018, version 1 of the entry of 5 December 2018.

In some embodiments the beta-lactamase may be a protein with 98 % or more identity to the sequence of SEQ ID NO: 6. It may for example be the beta-lactamase protein from *Escherichia coli* with the amino acid sequence of the SwissProt/UniProt

accession number R9URM7, version 1 of the sequence of 18 September 2013, version 19 of the entry of 5 December 2018. In some embodiments the beta-lactamase may be a protein with 96 % or more identity to the sequence of SEQ ID NO: 6. It may for example be the beta-lactamase protein from *Serratia marcescens* with the amino acid sequence of the SwissProt/UniProt accession number O33677, version 1 of the sequence of 1 January 1998, version 74 of the entry of 5 December 2018. In some embodiments the beta-lactamase may be a protein with 95 % or more identity to the sequence of SEQ ID NO: 6. It may for example be the beta-lactamase protein from *Acinetobacter baumannii* with the amino acid sequence of the SwissProt/UniProt accession number H9AXM0, version 1 of the sequence of 16 May 2012, version 29 of the entry of 5 December 2018.

In some embodiments the beta-lactamase may be a protein with 40 % or more identity to the sequence of SEQ ID NO: 6. As an illustrative example, the enzyme beta-lactamase from strain CDC 1551 of *Mycobacterium tuberculosis* having the sequence of the SwissProt/UniProt accession number P9WKD2, version 1 of the sequence of 16 April 2014, version 23 of the entry of 5 December 2018, has a sequence of 381 amino acids, of which 93 are identical to the sequence of SEQ ID NO: 6. In some embodiments the beta-lactamase may be a protein with 2 % or more identity to the sequence of SEQ ID NO: 6. As an illustrative example, the enzyme beta-lactamase from *Enterobacter cloacae* with the sequence of the SwissProt/UniProt accession number P05364, version 1 of the sequence of 1 November 1988, version 109 of the entry of 5 December 2018, has a sequence of 381 amino acids, of which 5 are identical to the sequence of SEQ ID NO: 6.

The vector disclosed herein furthermore includes a sequence that encodes an enzyme that confers resistance against the antibiotic puromycin. The respective enzyme may be aminoglycoside phosphotransferase (aph). The enzyme conferring resistance against the antibiotic puromycin may also be blasticidin S deaminase (bsd). The respective enzyme may also be puromycin-N-acetyltransferase (pac).

In some embodiments the sequence encoding the puromycin-N-acetyltransferase may be the sequence of bases 3094 to 3954 of cloning vector pL1F-3 of GenBank accession number KF955552.1, version 1 dated 3 March 2014. This sequence is SEQ ID NO: 7, and is also found e.g. as base positions 6391 to 7251 of the vector of SEQ ID NO: 87 of US patent US 10,113,179 or positions 6565 to 7425 of the vector of SEQ ID NO: 88 or positions 6979 to 7839 of the repair donor cassette of SEQ ID NO: 221 of the same patent. In some embodiments the sequence is the complementary sequence to SEQ ID NO: 7, which is SEQ ID NO: 8. In some embodiments the sequence encoding the puromycin-N-acetyltransferase may be the sequence of the extended spectrum beta lactamase (TEM) gene of *Escherichia coli* strain SRT41, which has GenBank accession number MG653169.1, version 1 dated 8 May 2018. Both sequences encode the amino acid sequence of SEQ ID NO: 6.

The promoter used for the enzyme that provides resistance to a bacterial host against ampicillin may be the beta-lactamase promoter, for example the promoter of the sequence of SEQ ID NO: 9. The complementary sequence is the sequence of SEQ ID NO: 10.

The vector includes an insertion site for a nucleotide sequence that encodes the target polypeptide to be expressed. The insertion site includes one or more restriction recognition sites for a restriction enzyme. Typically the vector has a multiple cloning site also called a polylinker. A multiple cloning site is a short segment of a nucleic acid sequence which contains a plurality of restriction sites. It may include up to about 20 restriction sites. A restriction site within a multiple cloning site is typically unique, occurring only once within a given plasmid. A multiple cloning site allows the insertion of a gene of interest into the region of the A multiple cloning site. The multiple cloning site generally immediately follows the promoter, and ends before the transcription terminator.

A target polypeptide expressed by a sequence that may be included into the insertion site may be any desired polypeptide. Two illustrative examples are an antibody, including a functional antibody fragment, and an enzyme. Two further illustrative examples are a growth factor and a blood coagulation factor.

The insertion site is operably linked to a CMV promoter. The CMV promoter may be of the sequence of SEQ ID NO: 4. The CMV promoter may also have a sequence that has 99 % or more identity to SEQ ID NO: 4. The CMV promoter may for example have or include the sequence of the hCMV immediate early promoter at positions 8186 to 8389 of the cloning vector pHR'-CMVLacZ of GenBank accession number AF105229.1, version 1 of the sequence as of 17 December 1998. As a further example, the CMV promoter may have or include the sequence of the CRU5 chimeric CMV promoter at positions 366 to 569 of the retroviral expression vector L149 of GenBank accession number EU753858.1, version 1 of the sequence as of 31 May 2009. The CMV promoter may in some embodiments have a sequence that has 98 % or more identity to SEQ ID NO: 4. It may for example have or include the sequence of the promoter at positions 8806 to 9009 of the Expression vector pDEST152 of GenBank accession number MH107058.1, version 1 of the sequence as of 30 October 2018. As another example, the CMV promoter may have or include the sequence of the hCMV-IE promoter at positions 1053 to 1256 of the mammalian expression vector pACCMVpLpA-E-hRac1-DN of GenBank accession number LT727056.1, version 1 of the sequence as of 6 February 2017.

The vector may include one or more further promoters operably linked to a restriction recognition site for a restriction enzyme, in order to be able to express a desired protein, the sequence for which may be inserted at the respective restriction site. As an example, a promoter of a T7 bacteriophage, such as a promoter of the T7 RNA Polymerase, may be included in the vector. The sequence of SEQ ID NO: 16 is an

example of a T7 promoter.

The bacterial replication origin of the vector may be chosen as desired. It is generally a plasmid origin. Typically the origin of replication is a high-copy number origin. In one embodiment the origin is the pMB1 origin. The bacterial origin of replication is in one embodiment the high-copy-number pUC origin. The pUC origin is the most common origin of replication that is used in bacterial vectors. It is derived from the pMB1 origin of the plasmid pBR322 but contains a point substitution within the origin as well as a deletion of the Rop/Rom gene. The letter 'p' indicates that the origin is of a circular, double-stranded DNA molecule, a plasmid. The letters 'UC' stand for "University of California".

In some embodiments the vector further contains an expression augmenting sequence element. Such elements have been disclosed in WO 1997/025420. A DNA sequence formed by ligating two sequences of SEQ ID NO: 1 of WO 1997/025420, namely nucleotides 8672 through 12273 and nucleotides 14290 through 14507 of SEQ ID NO: 1 of WO 1997/025420, has been described as having expression augmenting activity, see Working Example 7 of WO 1997/025420. The sequence of nucleotides 10100 through 14293 of SEQ ID NO: 1 of WO 1997/025420 was likewise disclosed as having expression augmenting activity. Aldrich, T.L., et al, Cytotechnology (1998) 28, 1-3, 9-17, disclose an EASE element termed a "truncated EASE element". A nucleic acid sequence corresponding to a respective DNA sequence may be used as an EASE in the context of the present disclosure.

The expression augmenting sequence element may in some embodiments be a portion of a known EASE. The EASE may in some embodiments include the sequence of SEQ ID NO: 13. The expression augmenting sequence element sequence of SEQ ID NO: 13 is for instance found in US patent US 6,596,514 in positions 8672 to 12273 of SEQ ID NO: 1 of the document, which has a length of 14507 bases. The EASE sequence of SEQ ID NO: 13 is for instance also included in the CHO sequence of GenBank accession number AF193761.1, version 1 as of 8 November 1998, which has a length of 14462 bases. In this sequence the EASE sequence of SEQ ID NO: 13 defines positions 8627 to 12228 of the sequence of the database entry. In some embodiments the EASE is defined by SEQ ID NO: 13. In some embodiments the EASE is defined by SEQ ID NO: 14.

Including the EASE into the vector that contains the PGK promoter, such as a PGK1 promoter – which is operably linked to the nucleotide sequence encoding a glutamine synthetase – leads to increased expression of the nucleotide sequence that encodes the target polypeptide, inserted in frame into the insertion site of the expression cassette. In some embodiments expression is transient expression. Including the EASE into the vector thus in some embodiments leads to high transient expression of the target polypeptide. In this regard the inventors have found that the ability of the EASE to promote stable integration leads to an extended state of the transient gene production

period of the vector, relative to the same vector lacking the EASE. As a longer period during which the plasmid is retained in the host cell, causing expression of the target polypeptide, leads to more target polypeptide being produced, the overall amount of target polypeptide produced is increased.

A target polypeptide that may be expressed by an expression vector according to the present disclosure may for example be a therapeutic polypeptide. Two illustrative examples of a therapeutic polypeptide are an adhesion molecule and a cytokine. Two further illustrative examples of a therapeutic polypeptide are an enzyme and a receptor. Another example of a therapeutic polypeptide is a lymphokine. Antibody light and/or heavy chains are yet further example of a therapeutic polypeptide. In some embodiments, the expression vector is adapted to express two target polypeptides, such as the individual polypeptide chains of a heterodimeric protein.

In some embodiments a vector as disclosed herein may contain two target polypeptide expression cassettes. These two target polypeptide expression cassettes may in some embodiments be identical in composition. Once sequences encoding the respective target polypeptide have been inserted into the vector, the target polypeptide expression cassettes in this case only differ in the sequences encoding the target polypeptides. In some embodiments two such target polypeptide expression cassettes are arranged in tandem in the vector.

A method as well as a use that employ an expression vector as disclosed herein may be a method or use, respectively, in which a polypeptide is being produced. Such a polypeptide is herein also termed the "target polypeptide". The method or use may be a method or use for achieving an improved yield of target protein. Such a method or use may involve culturing a recombinant host cell, such as a CHO or NS0 host cell which includes an expression vector as disclosed herein. In such an embodiment the expression vector generally includes a sequence that encodes the target polypeptide. As explained above, such a sequence is included at an insertion site for a nucleotide sequence encoding the target polypeptide. The respective sequence is typically in frame with the cytomegalovirus (CMV) promoter. In some embodiments the nucleotide sequence encoding the target polypeptide is in frame with the PGK promoter and the nucleotide sequence encoding a glutamine synthetase that are included in the selection cassette.

A respective method or use may involve culturing the host cell in a batch or in a fed-batch process. A respective method or use may also involve culturing the host cell in continuous mode. The cells are generally cultured under conditions allowing or promoting expression of the polypeptide. A respective method or use may include recovering the polypeptide.

A respective method or use may also include selecting the host cell on the basis of the presence of the selection cassette that contains the sequence encoding a glutamine synthetase, thereby leading to glutamine synthetase expression. The host cell may in

this regard be cultured in the absence of glutamine in the culture medium.

Any number of steps of a method according to the present disclosure, including the entire method, may be performed in an automated way – also repeatedly, using for instance commercially available robots. Computer executable instructions may for instance control data analysis or control mechanical courses of movements employed in a respective method.

The listing or discussion of a previously published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

The invention illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by exemplary embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

Other embodiments are within the appending claims. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

In order that the invention may be readily understood and put into practical effect, particular embodiments will now be described by way of the following non-limiting examples.

EXAMPLES

The examples illustrate the generation and use of expression vectors based on the vector pcDNA 3.1 (+).

Material

List of enzymes:

No.	Restriction Enzymes	Manufacturer	Catalog Number
1	NdeI	NEB	R0111S
2	NotI	NEB	R0189S
3	PvuI	NEB	R0150S
4	NruI	NEB	R0192S
5	EcoRV	NEB	R0195S
6	BglII	NEB	R0144S
7	BamHI	NEB	R0136S
8	SalI	NEB	R0138S
9	SphI	NEB	R0182S
10	SmaI	NEB	R0141S
11	SacI	NEB	R0156S

List of kits:

No.	Name of the Kit	Manufacturer	Catalog number
1	Qiagen quick gel extraction kit	Qiagen	K210012
2	Qiaprep Spin Mini prep kit	Qiagen	27106
3	Endofree-Maxi prep kit	Qiagen	12362

Reagents:

No.	Reagent	Manufacturer	Catalog number
1	Ethidium Bromide	SIGMA	E8751-1G
2	1kb Marker	Thermo Scientific	SM0311

Media Components Used:

No.	Media Name	Manufacturer	Catalog Number
1	LB Agar	Himedia	M557
2	LB Broth	Himedia	M575

Strains used:

No.	Strains and vectors	Company	Catalog Number
1	*E coli* Top-10	Invitrogen	NAP

Procedure

In order to assemble four vectors, three fragments were designed and synthesized from geneart named – GA Fragment 1: pMBL C-GS, GA Fragment 2: Puromycin gene fragment, and GA Fragment 3: EASE

Reference/Source for these three fragments is given below:

Source for GA fragment 1: pMBL C-GS

Table 1

No.	**vector component in pcDNA 3.1 (+)**	**Location (bp)**	**Modification for pMBL C-GS**
1	NAP	12	AG changed to AGATCTG to introduce BglII site
2	CMV promoter	232-819	No modification
3	Multiple cloning site	863-882	Modification to include - NotI, XhoI, and BamHI site
4	BGH polyadenylation sequence	1028-1252	It was removed and replaced with sv40 pA from pCHO1.0
5	f1 origin	1298-1726	The sequence from 1298-3234 was deleted
6	sv40 early promoter and origin	1731-2074	

No.	vector component in pcDNA 3.1 (+)	Location (bp)	Modification for pMBL C-GS
7	Neomycin resistance gene (ORF)	2136-2930	and replaced with PGK-GS selection cassette
8	sv40 early polyadenylation signal	3104-3234	No modification
9	pUC origin	3617-4287	No modification
10	Ampicillin resistance gene	4432-5428	No modification
11	SalI restriction site	5427	C changed to A, to remove SalI site

A vector map of pcDNA 3.1 (+), which is a parent vector for the GA fragment 1, is shown in Fig. 1A. The modifications were made using Vector NTI Version 11.5.1 to parent in-house vector. The in-house vector map of resultant vector pMBL C-GS is shown in Fig. 1B. The specific modifications are captured in the table above.

Reference for Puromycin (GA Fragment 2):

The puromycin gene cassette was sourced from the pCHO1.0 vector from Invitrogen. The gene for puromycin was synthesized from GA with a SalI site at the 5' and SmaI site at the 3' end of the sequence.

Reference for EASE (GA Fragment 3):

The EASE fragment was sourced from GenBank accession number AF193761.1 - 8672 bp to 12274 bp. The sequence was ordered from Geneart with a BglII site at the 5' end and a BamHI site at the 3' end of sequence.

Process workflow

The three fragments were synthesized from geneart and reference for each is given in Table 1 above:

1. GA Fragment 1: pMBL C-GS
2. GA Fragment 2: Puromycin gene fragment (labelled as 16AA4TQC_PURO__pMK-RQ)
3. GA Fragment 3: EASE (labelled as 16AASA2C_Frag_3_pMK from GA)



Reconstitution of lyophilized GA fragments, propagation, plasmid isolation, and glycerol stock preparation



The following four in-house vectors were assembled:

1. pMBL CE-GS
2. pMBL CE-Puro
3. pMBL C-Puro
4. pMBL C-GS (Received from geneart)



Creation of pMBL CE-GS:

Cloning of EASE (GA Fragment 3) at BglII site in pMBL C-GS



Creation of pMBL CE-Puro:

Cloning of puromycin gene (GA Fragment 2) at SalI-SmaI site in pMBL CE-GS



Creation of pMBL C-Puro:

Cloning of puromycin gene (GA Fragment 2) at SalI-SmaI site in pMBL C-GS

The following Table 2 provides a summary on the assembled vectors.

Table 2:

Nomenclature	EASE	hCMV promoter	MCS	SV40 pA	PGK-GS	SV40 pA	PUC origin	Ampicillin marker	Puromycin resistant gene	Total size (in bp)

pMBL CE-GS	Y	Y	Y	Y	Y	Y	Y	Y	N	8974
pMBL CE-Puro	Y	Y	Y	Y	N	N	Y	Y	Y	8331
pMBL C-Puro	N	Y	Y	Y	N	N	Y	Y	Y	4723
pMBL C-GS	N	Y	Y	Y	Y	Y	Y	Y	N	5366

No.	Nomenclature	pMBL CE-GS	pMBL CE-Puro	pMBL C-Puro	pMBL C-GS
1	EASE	Yes	Yes	No	No
2	hCMV promoter	Yes	Yes	Yes	Yes
3	MCS	Yes	Yes	Yes	Yes
4	SV40 pA	Yes	Yes	Yes	Yes
5	PGK-GS	Yes	No	No	Yes
6	SV40 pA	Yes	No	No	Yes
7	PUC origin	Yes	Yes	Yes	Yes
8	Ampicillin marker	Yes	Yes	Yes	Yes
9	Puromycin resistance gene	No	Yes	Yes	No
	Total size (in bp)	8974	8331	4723	5366

Characterization of vector pMBL C-GS

The vector pMBL C-GS was obtained from Geneart (GA Fragment 1) was re-suspended in 25µL of sterile water to obtain a final concentration of 200 ng/µL.

1µL was transformed to CCB00024/CHE chemically competent cells. A single colony was streaked and inoculated for plasmid isolation. The overnight culture was then stored as a glycerol stock. The plasmid was isolated.

The plasmid stock was analysed by restriction digestion and subsequent agarose gel electrophoresis. Fig. 5A depicts the obtained gel.

Characterization of vector pMBL CE-GS

The EASE fragment was obtained from Geneart (GA Fragment 3). The EASE element had been sub cloned upstream of CMV promoter at the BglII site, thereby obtaining vector pMBL C-GS. The EASE fragment was resuspended in 50µL of sterile water to obtain a final concentration of 100 ng/µL.

1µL was transformed to CCB00024 chemically competent cells.

A single colony obtained was streaked and the plasmid was isolated. The EASE fragment was released from the GA Fragment 3 by using the BglII+BamHI sites, and cloned in pMBL C-GS (PLB00876) at the BglII site. The resulting vector obtained is pMBL CE-GS.

The identity of vector pMBL CE-GS was analysed by restriction digestion and subsequent agarose gel electrophoresis. Fig. 5B depicts the obtained gel.

Characterization of vector pMBL CE-Puro

The puromycin gene (GA Fragment 2) was obtained from Geneart. The EASE element had been sub cloned upstream of the CMV promoter at the BglII site, and the GS ORF had been replaced with a Puromycin selection marker for CHO at the SmaI/SalI sites, supra. The fragment was resuspended in 50µL of sterile water to obtain a final concentration of 100 ng/µL.

1µLwas transformed to CCB00024 chemically competent cells. A single colony was streaked and the plasmid was isolated.

The puromycin gene from GA fragment 2 was cloned into the vector pMBL CE-GS at the SmaI+SalI site. The resulting vector obtained is the vector pMBL CE-Puro.

The identity of vector PMBL CE-Puro was analysed by restriction digestion and subsequent agarose gel electrophoresis. Fig. 5C depicts the obtained gel.

Characterization of vector pMBL C-Puro

The puromycin gene (GA Fragment 2) was obtained from Geneart. The GS ORF had been replaced with a Puromycin selection marker for CHO at the SmaI/SalI sites, supra. The fragment was resuspended in 50µL of sterile water to obtain a final concentration of 100 ng/µL.

1µLwas transformed to CCB00024/CHE chemically competent cells. A single colony was streaked and the plasmid was isolated.

The puromycin fragment from GA fragment 2 was released by using SmaI+SalI sites and cloned into vector pMBL C-GS at the SmaI+SalI sites. The resulting vector obtained is the vector pMBL C-Puro.

The identity of vector PMBL C-Puro was analysed by restriction digestion and subsequent agarose gel electrophoresis. Fig. 5D depicts the obtained gel.

Vector evaluation

The assembled vectors were further evaluated on the basis of their capability to drive heterologous gene expression in a CHO based cell line. As a start, pMBL CE-GS was selected for evaluation. The following constructs were created in pMBL CE-GS:

Experiment	Experiment 1	Experiment 2		Experiment 3			
Construct tested	GFP/pMBL CE-GS	Bmab700/ pMBL CE-GS	Bmab700/ pMBL C-GS	Fmab25	Fmab26	Fmab27	Fmab28
	Transfection efficiency control	Titer comparision for Biosimilar mAb		Titer comparision for fusion mAb			

Table 3: Constructs tested for vector evaluation

Experiment 1:

GFP was cloned into the vector pMBL CE-GS at XhoI – NotI sites. This plasmid acts as a transfection efficiency control to evaluate pMBL CE-GS. The plasmid was evaluated by transient gene expression. Transient transfections were performed in co-transfection mode. All transfections were carried out by using the ExpiCHO™ Expression System (Cat. no. A29133 from Thermo Fisher Scientific Inc., Waltham, MA USA), which is a high-yield transient expression system based on suspension-adapted Chinese Hamster Ovary (CHO) cells. ExpiCHO-S cells were thawed as per the expiCHO-S expression system manual. After a minimum of two passages post thaw, the cells were seeded for transfection at 3.5 x 10^6 cells/ml one day before transfection. On day 1 post seeding, the cell count was estimated and adjusted to 6.0 x 10^6 cells/ml by diluting with fresh, pre-warmed medium for each transfection. Transfection was carried out according to the expiCHO-S expression system manual. DNA was transfected at 1µg/ml final concentration in culture. Expifectamine and feeds were added according to the manufacturer's protocol. Max titre protocol was followed for all transfections. Transfection flasks were cultivated at ~37° C in an incubator with a humidified atmosphere of 8% CO_2 in air on an orbital shaker. VCD and viabilities were estimated on days, 0, 2, 5, 7, 9, and 12. Cultures were harvested

when viability dropped below 50%.

Experiment 2:

Heavy chain and light chain of an anti-CTLA4 antibody (Bmab700) were cloned into the vectors pMBL CE-GS and pMBL C-GS at the XhoI - NotI site. The resulting constructs were evaluated in comparison to Bmab700-SSC/pCHO1.0 M (HC and LC cloned in two individual vectors and co-transfected) and Bmab700-DGC/pCHO1.0 (HC and LC cloned in a single vector). The four constructs were compared by transient gene expression.

Transient transfections were performed in co-transfection mode. All transfections were carried out by using the ExpiCHO™ Expression System (Cat. no. A29133 from Thermo Fisher Scientific Inc., Waltham, MA USA), a high-yield transient expression system based on suspension-adapted CHO cells. ExpiCHO-S cells were thawed according to the expiCHO-S expression system manual. After a minimum of two passages post thaw, the cells were seeded for transfection at 3.5 x 10^6 cells/ml one day before transfection. On day 1 post seeding, the cell count was estimated and adjusted to 6.0 x 10^6 cells/ml by diluting with fresh, pre-warmed medium for each transfection. Transfection was carried out as per expiCHO-S expression system manual. DNA was transfected at 1µg/ml final concentration in culture. Expifectamine and feeds were added according to the manufacturer's protocol. Max titre protocol was followed for all transfections. Transfection flasks were cultivated at ~37° C incubator with a humidified atmosphere of 8% CO_2 in air on an orbital shaker. VCD and viabilities were estimated on days, 0, 2, 5, 7, 9, and 12. Cultures were harvested on day-15 and results are depicted in Fig.7A.

Culturing was carried out in shake flasks of 125 ml capacity, in which 25 ml running volume was used.

Results are depicted in Fig. 7A.

Experiment 3:

HC and LC of anti-CD20 fusion mabs were cloned into the vector pMBL CE-GS at the XhoI - NotI site. The resulting constructs were evaluated for titer analysis by transient gene expression.Transient transfections were performed in co-transfection mode. All transfections were carried out by using The ExpiCHO™ Expression System (Cat. no. A29133 from Thermo Fisher Scientific Inc., Waltham, MA USA), a high-yield transient expression system based on suspension-adapted CHO cells. ExpiCHO-S cells were thawed as per the expiCHO-S expression system manual. After a minimum of 2 passages post thaw, the cells were seeded for transfection at 3.5 x 10^6 cells/ml one day before transfection. On day 1 post seeding, the cell count was estimated and adjusted to 6.0 x 10^6 cells/ml by diluting with fresh, pre-warmed medium for each transfection. Transfection was carried out as per expiCHO-S expression system manual. DNA was transfected at 1µg/ml final concentration in

culture. Expifectamine and feeds were added as per manufacturer's protocol. The max titre protocol was followed for all transfections. Transfection flasks were cultivated at ~37° C incubator with a humidified atmosphere of 8% CO_2 in air on an orbital shaker. VCD and viabilities were estimated on days, 0, 2, 5, 7, 9, and 12. Cultures were harvested on day14. Results are depicted in Fig. 7B.

WE CLAIM:

1. An expression vector for mammalian cells, comprising

   (a) a selection cassette comprising a nucleotide sequence encoding a glutamine synthetase as a eukaryotic selection marker, operably linked to a 3-phosphoglycerate kinase (PGK) promoter and a polyadenylation (pA) signal;

   (b) a selection cassette comprising a nucleotide sequence encoding an enzyme that confers resistance against an antibiotic to a bacterial host as a bacterial selection marker, operably linked to a suitable promoter;

   (c) an expression cassette for a target polypeptide, comprising an insertion site for a nucleotide sequence encoding the target polypeptide, operably linked to a cytomegalovirus (CMV) promoter and a polyadenylation (pA) signal; and

   (d) a bacterial origin of replication.

2. The expression vector of claim 1, further comprising an expression augmenting sequence element (EASE).

3. The expression vector of claim 1 or 2, further comprising a selection cassette comprising a nucleotide sequence encoding an enzyme that confers resistance against the antibiotic puromycin as a eukaryotic selection marker, operably linked to a 3-phosphoglycerate kinase (PGK) promoter and a polyadenylation (pA) signal.

4. The expression vector of claim 3, wherein the enzyme that confers resistance against the antibiotic puromycin is puromycin-N-acetyltransferase (pac).

5. The expression vector of any one of the preceding claims, wherein the antibiotic against which the bacterial selection marker confers resistance is Ampicillin, and wherein the enzyme that confers resistance against Ampicillin to a bacterial host is beta-lactamase.

6. The expression vector of any one of the preceding claims, wherein the bacterial origin of replication is the pUC origin of replication.

7. The expression vector of any one of the preceding claims, wherein the glutamine synthetase as a eukaryotic selection marker has a sequence of at least 90 % identity to the sequence of SEQ ID NO: 3 encoding a glutamine synthetase capable of catalyzing the ATP-dependent conversion of glutamate and ammonia to glutamine, or wherein the glutamine synthetase as a eukaryotic selection marker is a mammalian glutamine synthetase.

8. The expression vector of any one of the preceding claims, wherein the glutamine synthetase has an amino acid sequence of at least 96 % identity to the sequence

of SEQ ID NO: 5 and is capable of catalyzing the ATP-dependent conversion of glutamate and ammonia to glutamine.

9. The expression vector of any one of the preceding claims, wherein the 3-phosphoglycerate kinase (PGK) promoter has a sequence of at least 98 % identity to the sequence of SEQ ID NO: 2 and is a functional promoter, and/or wherein the glutamine synthetase as a eukaryotic selection marker is a CHO glutamine synthetase having at least 97 % identity to the sequence of SEQ ID NO: 3 encoding a glutamine synthetase capable of catalyzing the ATP-dependent conversion of glutamate and ammonia to glutamine.

10. The expression vector of any one of the preceding claims, wherein the pA signal is the simian virus 40 pA signal.

11. The expression vector of any one of the preceding claims, being comprised in a recombinant CHO or NS0 host cell.

12. A method of producing a polypeptide, the method comprising culturing the recombinant CHO or NS0 host cell according to claim 11 under conditions suitable for expressing a heterologous target polypeptide, wherein the insertion site of the expression cassette for a target polypeptide comprises a nucleotide sequence encoding the polypeptide as a heterologous target polypeptide.

13. The method of claim 12, wherein the polypeptide comprises an antibody or a fragment thereof.

14. The use of an expression vector according to any one of claims 1 to 11 for expressing a polypeptide at high levels.

15. The use of claim 14, wherein the polypeptide comprises an antibody or a fragment thereof.