TRANSCRIPTIONAL REGULATORY ELEMENT AND ITS USE IN ENHANCING THE EXPRESSION OF HETEROLOGOUS PROTEIN

TECHNICAL FIELD

The present disclosure relates to the fields of molecular biology and bioengineering. In particular, the present disclosure relates to a novel transcriptional regulatory element which uses mammalian cells to express a heterologous protein. Specifically, the present disclosure relates to a transcriptional regulatory element WXRE (WuXi Regulatory Element) which is used in a eukaryotic cell line to prepare heterologous protein and enhance the expression level of the above-mentioned protein, and an expression system of the heterologous protein comprising WXRE, as well as the use of the above-mentioned expression system in producing heterologous protein.

BACKGROUND

In biological studies, the study of the proteins has received more and more attention, and the most important thing for the protein research is the selection of a protein expression system. The protein expression system refers to a molecular biological technology which uses model organisms such as bacteria, yeast, plant cells or animal cells to express heterologous proteins. The common protein expression systems are divided into prokaryotic expression systems and eukaryotic expression systems.

Among them, the prokaryotic expression system is a system which obtains heterologous proteins by prokaryotes and mainly includes *Escherichia coli* expression system, *Bacillus subtilis* expression system, Streptomycin expression system, and the like. Among them, the *Escherichia coli* expression system is the most widely used. The characteristics of the prokaryotic expression system are rapid growth of the host bacteria, easy cultivation, convenient operation, low price, clear genetic background, safe genes and high protein expression level. However, the prokaryotic expression system may not regulate the expression time and the expression level; meanwhile, the expression products of the prokaryotic expression system may exist in the form of an inclusion body with low biological activity, and the post-translational processing and modifying system is imperfect (for example,

glycosylation modification may not be performed).

On the other hand, the eukaryotic expression system mainly includes the expression systems of mammalian cells, yeast cells and insect cells, which are commonly used methods to express heterologous proteins in recent years. It supplements some functions which are deficient in the prokaryotic expression systems. For example, stable disulfide bonds can be formed with the eukaryotic expression system, and after a protein is translated, the protein can be correctly modified, which enables the expressed protein to have more natural activity instead of being degraded or forming inclusion bodies. Among them, a mammalian expression system has the characteristics of being capable of inducing the highly efficient expression, performing correct folding of the expressed proteins and performing complex glycosylation modification accurately, having protein activity which is close to that of the natural protein, having no need to remove the endotoxins, and the like. Meanwhile, the mammalian expression system is the only system that can express complex proteins. As for the production of antibodies such as the humanized monoclonal antibodies, said mammalian expression system has the characteristics of being able to be produced in large amount, good humanization and the like. The host cells commonly used in the mammalian expression systems include CHO cells, COS cells, BHK cells and the like.

Adalimumab (trade name: Humira) is an anti-human tumor necrosis factor (TNF) humanized monoclonal antibody, which is approved by NMPA (National Medical Products Administration) for the treatment of rheumatoid arthritis and ankylosing spondylitis and has good therapeutic effect.

Granulocyte-macrophage colony stimulating factor (GM-CSF) is a hematopoietic cell growth factor having pleiotropy and acts to regulate in the developmental and mature stages of granulocytic series and macrophages. GM-CSF also plays a key role in the differentiation of the monocytes in synovial tissues into the inflammatory dendritic cells and involves in inducing and maintaining the emergence and the development of acute arthritis. Meanwhile, the tests of serum and synovium are performed for patients suffering from rheumatoid arthritis. It is found that the level of GM-CSF factor increases significantly and is closely related to disease activity, which is mainly reflected in the deterioration of bone erosion as well as the synovial lining cells and the underlayer cells which are infiltrated with large amount of

macrophages.

PD-1 is a key immune checkpoint receptor expressed by the activated T and B cells, and mediates the immunosuppression. PD-1 is a member of the CD28 receptor family, this family includes CD28, CTLA-4, ICOS, PD-1 and BTLA. Two glycoprotein ligands of PD-1 on cell surface, i.e., programmed death ligand-1 (PD-L1) and programmed death ligand-2 (PD-L2), have been identified. They are expressed on the antigen-presenting cells as well as in a variety of human cancers, and they have been shown to downregulate the activation of T cells and the cytokine secretion upon binding to PD-1.

PD-L1 is also referred to as CD274 or B7-H1. PD-L1 has up-regulated expression in a variety of tumor cells, binds to the PD-1 on T cells, inhibits the proliferation and activation of T cells, makes T cells in an inactivated state, and eventually induces the immune escape. The inhibitors of PD-L1 can block the binding of PD-1 and PD-L1, up-regulate the growth and proliferation of T cells, enhance the recognition of T cells to tumor cells, activate its attacking and killing functions, and achieve the anti-tumor effect by mobilizing the immune function of human body itself.

Since monoclonal antibodies such as Adalimumab, the monoclonal antibodies against PD-1, and the monoclonal antibodies against PD-L1 can all be produced by the mammalian expression systems, and the problem that the expression level of the heterologous proteins is low exists in the mammalian expression systems, there is an urgent need to improve the mammalian expression systems known in the prior art and provide a system which is capable of achieving significantly higher expression level of heterologous proteins. Meanwhile, the above-mentioned systems are needed to produce the humanized monoclonal antibodies.

SUMMARY

Technical Problem

The present disclosure provides a DNA sequence which is derived from CHO (Chinese Hamster Ovary) cells and can be used to enhance the expression of the heterologous protein, said DNA sequence is capable of being used for enhancing expression level of the heterologous proteins of the mammalian expression systems.

Solution to Problem

The technical solutions related in the present disclosure are as follows.

(1). A polynucleotide, wherein the polynucleotide is selected from any one of (i) to (iv):

(i) a nucleotide sequence comprising a sequence as shown in any sequence of SEQ ID NOs:3-9;

(ii) a nucleotide sequence comprising a reverse complementary sequence of the sequence as shown in any sequence of SEQ ID NOs:3-9;

(iii) a reverse complementary sequence of a sequence capable of hybridizing with the nucleotide sequence as shown in (i) or (ii) under a high stringency hybridization condition or a very high stringency hybridization condition;

(iv) a sequence having at least 90% sequence identity, alternatively at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity, most preferably at least 99% sequence identity with the nucleotide sequence as shown in (i) or (ii).

(2) The polynucleotide according to (1), wherein the polynucleotide also comprises any one selected from (v) to (viii):

(v) a nucleotide sequence comprising a sequence as shown in SEQ ID NO: 13;

(vi) a nucleotide sequence comprising a reverse complementary sequence of the sequence as shown in SEQ ID NO:13;

(vii) a reverse complementary sequence of a sequence capable of hybridizing with the nucleotide sequence as shown in (v) or (vi) under a high stringency hybridization condition or a very high stringency hybridization condition;

(viii) a sequence having at least 90% sequence identity, alternatively at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity, most preferably at least 99% sequence identity with the nucleotide sequence as shown in (v) or (vi).

(3) The polynucleotide according to (1) or (2), wherein the polynucleotide is capable of increasing the protein expression level of a protein expression system; alternatively, the protein expression system is selected from the protein expression systems of eukaryotic cells; preferably, the protein expression systems of eukaryotic cells are selected from the protein expression systems of CHO (Chinese Hamster Ovary) cells.

(4) The polynucleotide according to (1) or (2), wherein the polynucleotide is selected from a sequence comprising the sequence as shown in SEQ ID NO:4, or a sequence comprising the reverse complementary sequence of the sequence as shown in SEQ ID NO:4; alternatively, the polynucleotide is selected from a nucleotide sequence comprising the sequence as shown in SEQ ID NO:4 and the sequence as shown in SEQ ID NO:13, or a nucleotide sequence comprising the reverse complementary sequence of the sequence as shown in SEQ ID NO:4 and the reverse complementary sequence of the sequence as shown in SEQ ID NO:13.

(5) The polynucleotide according to (3), wherein the protein expression system is used to express an antibody, a fusion protein or a recombinant protein; alternatively, the antibody is selected from monoclonal antibodies, the fusion protein is the antibody against programmed death-1 (PD-1) or the antibody against programmed death-ligand 1 (PD-L1); preferably, the monoclonal antibody is Adalimumab, and said antibody against PD-1 is pembrolizumab.

(6) A WXRE transcriptional regulatory element used for enhancing the protein expression level of a protein expression system, wherein the WXRE transcriptional regulatory element comprises the polynucleotide according to any one of (1) to (5).

(7) A protein expression vector or a protein expression system, wherein the protein expression vector or the protein expression system comprises the polynucleotide according to any one of (1) to (5) or the WXRE transcriptional regulatory element according to (6); alternatively, the protein expression vector or the protein expression system also comprises at least a promoter and at least a restriction enzyme site; preferably, the protein expression vector or the protein expression system is selected from the protein expression vectors of mammalian cells or the protein expression systems of mammalian cells; more preferably, the mammalian cells are CHO cells.

(8) A cell line, wherein the cell line comprises the protein expression vector or the protein expression system according to (7).

(9) A kit, wherein the kit comprises the protein expression vector or the protein expression system according to (7), or the cell line according to (8).

(10) Use of the protein expression vector or the protein expression system according to (7), or the cell line according to (8) in the preparation of a reagent or a kit for detecting an animal disease due to the abnormality of protein expression.

(11) The use according to (10), wherein the animal disease is capable of causing the abnormal expression of a target protein in the animal, and the protein expression vector, or the protein expression system, or the cell line is capable of secreting an antibody of the target protein; alternatively, the animal is selected from mammals; preferably, the mammal is human.

In a technical solution, it can be known from the prior art that, as for rheumatoid arthritis, it is found that the level of GM-CSF factor increases significantly and is closely related to disease activity, mainly reflected in the deterioration of bone erosion as well as the synovial lining cells and the under layer thereof which are infiltrated with large amount of macrophages. Therefore, an antibody of GM-CSF can be prepared and obtained via the protein expression vector or the protein expression system related in the present disclosure, wherein the sequence of the antibody of GM-CSF is known in the prior art. Illustratively, the sequence of the antibody of GM-CSF can be selected from the sequences as shown in WO 2018050111 A1. After the antibody of GM-CSF is obtained, it is used to diagnose rheumatoid arthritis by further detecting the expression level of GM-CSF.

(12) Use of a protein expressed by the protein expression vector or the protein expression system according to (7) or the cell line according to (8) in the preparation of a drug for treating or preventing an animal disease; alternatively, the animal disease is selected from tumors or autoimmune diseases.

(13) The use according to (12), wherein the tumor is selected from cancers, and/or the autoimmune disease is selected from rheumatoid arthritis or ankylosing spondylitis.

(14) A method for preparing a protein, wherein the protein expression vector or the protein expression system according to (7), or the cell line according to (8) is selected to secrete the protein; alternatively, the protein is selected from an antibody, a fusion protein or a recombinant protein; preferably, the antibody is selected from monoclonal antibodies, the fusion protein is the antibody against programmed death-1 (PD-1) or the antibody against programmed death-ligand 1 (PD-L1); more preferably, the monoclonal antibody is Adalimumab, and said antibody against PD-1 is pembrolizumab.

(15) A screening method of a stable cell line highly expressing target protein, wherein a target gene which encodes the target protein is transfected into the cell by using the protein expression vector or the protein expression system according to (7), and the stable cell line

highly expressing target protein is screened and obtained; alternatively, the protein expression vector or the protein expression system is selected from the protein expression vectors of mammalian cells or the protein expression systems of mammalian cells; more preferably, the mammalian cells are CHO cells.

(16) The screening method according to (15), wherein the screening method includes antibiotic screenings aiming at selection markers or drug pressure screenings aiming at amplified marker genes.

(17) A cell line, which is obtained by using the screening method according to any one of (15) to (16).

(18) A method for detecting an animal disease, wherein a protein secreted by the protein expression vector or the protein expression system according to (7) or the cell line according to (8) is adopted to detect whether an animal suffers from the disease; wherein the animal disease is capable of causing the abnormal expression of a target protein in the animal; and,

(i) if the secreted protein is capable of interacting with the target protein, it indicates that the animal suffers from the disease;

(ii) if the secreted protein is capable of not interacting with the target protein, it indicates that the animal does not suffer from the disease.

(19) The method according to (18), wherein the protein expression vector, or the protein expression system, or the cell line is capable of secreting an antibody of the target protein; alternatively, the animal is selected from mammals; preferably, the mammal is human.

(20) A method of treating or preventing an animal disease, wherein a protein secreted by the protein expression vector or the protein expression system according to (7) or the cell line according to (8) is administered to the animal; alternatively, the animal disease is selected from tumors or autoimmune diseases.

(21) The method according to (20), wherein the tumor is selected from cancers, and/or the autoimmune disease is selected from rheumatoid arthritis or ankylosing spondylitis.

(22) A method of preparing a target antibody, wherein the method includes the following steps:

(i) by the method for preparing a protein according to (14), preparing and obtaining the protein;

(ii) using the protein obtained in step (i) to immunize an animal, so as to obtain the corresponding target antibody.

(23) A vector, wherein the vector comprises at least any one of the polynucleotide according to (1) or (2).

(24) The vector according to (23), wherein the vector also comprises a CMV promoter; alternatively, the CMV promoter comprises a sequence as shown in SEQ ID NO: 16, or the CMV promoter comprises a sequence having at least 90% sequence identity, alternatively at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity, most preferably at least 99% sequence identity with the sequence as shown in SEQ ID NO:16.

(25) The vector according to (23) or (24), wherein the vector comprises one or more genes encoding one or more target proteins.

(26) The vector according to (25), wherein the target protein is selected from a group consisting of an antibody, a fusion protein, an enzyme, a soluble protein, a membrane protein, a structural protein, a ribosome protein, a zymogen, a cell surface receptor protein, a transcriptional regulatory protein, a translational regulatory protein, a chromatin protein, a hormone, a cell cycle regulatory protein, a G protein, a neuroactive peptide, an immunomodulatory protein, a blood component protein, an ion gate protein, a heat shock protein, dihydrofolate reductase, an antibiotic resistance protein, a functional fragment of any one of the target proteins, an epitope fragment of any one of the target proteins, and any combination thereof.

(27) An isolated host cell, wherein the host cell comprises the vector according to any one of (23) to (26).

(28) A method of preparing a host cell which expresses a target protein stably, wherein the method comprises a step of transforming an initial host cell by using the vector according to any one of (23) to (26).

(29) The host cell according to (27) or the method according to (28), wherein the host cell is a Chinese hamster ovary cell.

(30) A method of preparing a target protein, wherein the method comprises preparing the target protein by using the host cell according to (27) or via the method according to (28) or

(29).

Alternatively, the present invention provides the following embodiments.

(31). A polynucleotide comprising a nucleotide sequence that is at least 85% identical to a regulatory sequence selected from the group consisting of SEQ ID NOs:3-9, a promotor, and a heterologous sequence that encodes a polypeptide, wherein the regulatory sequence, the promotor, and the heterologous sequence that encodes the polypeptide are operably linked together.

(32). The polynucleotide of claim (31), wherein the polynucleotide further comprises an EF1αI gene intron.

(33). The polynucleotide of claim (32), wherein the EF1αI gene intron comprises a sequence that is at least 80% identical to SEQ ID NO: 13.

(34). The polynucleotide of claim (33), wherein the regulatory sequence selected from the group consisting of SEQ ID NOs: 3-9 has a forward direction.

(35). The polynucleotide of claim (34), wherein the regulatory sequence selected from the group consisting of SEQ ID NOs: 3-9 has a reverse direction.

(36). A host cell for enhanced expression of an antibody, the cell comprising

a first polynucleotide encoding an antibody heavy chain or fragment thereof, wherein the first polynucleotide is operably linked to a sequence that is at least 85% identical to a sequence of any of SEQ ID NOs:3-9 or a reverse complementary sequence of the sequence of any of SEQ ID NOs:3-9; and

a second polynucleotide encoding an antibody light chain or fragment thereof, wherein the second polynucleotide is operably linked to a sequence that is at least 85% identical to a sequence of any of SEQ ID NOs:3-9 or a reverse complementary sequence of the sequence of any of SEQ ID NOs:3-9.

(37). The host cell of claim (36), wherein the first polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 4 and the second polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 4.

(38). The host cell of claim (36), wherein the first polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 4 and the second polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 9.

(39). The host cell of claim (36), wherein the first polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 9 and the second polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 9.

(40). The host cell of claim (36), wherein the first polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 4 and the second polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 17.

(41). The host cell of claim (36), wherein the first polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 9 and the second polynucleotide is operably linked to a sequence that is at least 85% identical to SEQ ID NO: 17.

(42). An expression vector comprising an expression cassette, wherein the expression cassette comprises a promoter operably linked to a nucleic acid sequence encoding a polypeptide, and a regulatory sequence that is at least 85% identical to a sequence selected from the group consisting of SEQ ID NOs:3-9, wherein the regulatory sequence is operably linked to the promotor.

(43). The expression vector of claim (42), wherein the expression vector has a sense strand and an anti-sense strand, and the sense strand comprises a sequence (from 5' to 3') that is at least 85% identical to a sequence selected from the group consisting of SEQ ID NOs:3-9.

(44). The expression vector of claim (42), wherein the expression vector has a sense strand and an anti-sense strand, and the anti-sense strand comprises a sequence (from 5' to 3') that is at least 85% identical to a sequence selected from the group consisting of SEQ ID NOs:3-9.

Effects of Invention

In a technical solution, the use of the transcriptional regulatory element WXRE listed in the present disclosure can greatly increase the expression amount of a heterologous protein and has great contribution to the production of mammalian proteins.

In some technical solutions, the expression amount of a heterologous protein is increased by about or at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%. In some technical solutions, the expression amount of a heterologous protein is increased up to 80%.

In another technical solution, the use of the transcriptional regulatory element WXRE listed in the present disclosure can enable a heterologous protein to still maintain its biological

activity while the expression level is greatly increased.

In another technical solution, the transcriptional regulatory element WXRE listed in the present disclosure is able to be used together with other transcriptional regulatory elements as a whole, and maintains its biochemical activity while enabling the expression level of the heterologous protein to increase greatly.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a schematic diagram of a GFP-expressing vector without WXRE inserted therein.

Figure 2 illustrates a schematic diagram of a GFP-expressing vector with WXRE inserted therein.

Figure 3 illustrates the influence on the expression amount of the fusion protein after adding transcriptional regulatory element A~G, wherein A1 and A2 illustrate the forward and reverse directions of transcriptional regulatory element A respectively, and so forth.

Figure 4 illustrates the influence on the specific productivity of the expression of the fusion protein after adding transcriptional regulatory element A~G, wherein A1 and A2 illustrate the forward and reverse directions of transcriptional regulatory element A respectively, and so forth.

Figure 5 illustrates a schematic diagram of a vector which expresses the heavy chain of Adalimumab and has WXRE inserted therein, wherein HC means the heavy chain.

Figure 6 illustrates a schematic diagram of a vector which expresses the light chain of Adalimumab and has WXRE inserted therein, wherein LC means the light chain.

Figure 7 illustrates a comparison of the expression amount of Adalimumab on 14^{th} day under different combined conditions of the transcriptional regulatory elements, wherein in sample 1 to sample 6, the components of the transcriptional regulatory element in the upstream of the heavy chain and the transcriptional regulatory element in the upstream of the light chain are found in Table 6.

Figure 8 illustrates a schematic diagram of a vector which expresses the heavy chain of an antibody and comprises WXRE and EF1αI intron (i.e. the sequence as shown in SEQ ID NO: 13) inserted therein, wherein HC means the heavy chain.

Figure 9 illustrates a schematic diagram of a vector which expresses the light chain of an antibody and comprises WXRE and EF1αI intron (i.e., the sequence as shown in SEQ ID NO: 13) inserted therein, wherein LC means the light chain.

Figure 10 illustrates a schematic diagram of a vector which expresses an antibody heavy chain.

Figure 11 illustrates a schematic diagram of a vector which expresses an antibody light chain.

DETAILED DESCRIPTION

Definition

When used in the claims and/or the specification, the word "a" or "an" or "the" may refer to "one", but may also refer to "one or more", "at least one" and "one or more than one".

As used in the claims and the specification, the words "comprising", "having", "including" or "containing" means inclusive or open-ended, and does not exclude additional, unrecited elements or method steps. Meanwhile, "comprising", "having", "including" or "containing" may also mean closed-ended, and excludes additional, unrecited elements or method steps.

Throughout this application, the term "about" denotes that a value includes the standard deviation of the error of the device or method used to determine this value.

Although the content disclosed supports that the definition of the term "or" is a substitute only as well as "and/or". Unless it is clearly denoted that it is only a substitute or the substitutes are mutually exclusive, the term "or" in the claims refers to "and/or".

The "transcriptional regulatory element" in the present disclosure refers to some polynucleotides involved in the regulation of gene transcription. Alternatively, the above-mentioned polynucleotide may be selected from DNA(s), which mainly include a promoter, an enhancer, an insulator, and the like. In the present disclosure, the transcriptional regulatory element is also referred to as WXRE (WuXi Regulatory Element). Illustratively, The WXREs in the present disclosure include transcriptional regulatory element A, transcriptional regulatory element B, transcriptional regulatory element C, transcriptional regulatory element D, transcriptional regulatory element E, transcriptional regulatory element F, and transcriptional regulatory element G.

The meaning of the "reverse complementary sequence" in the present disclosure is a sequence which is opposite to the direction of the sequence of the original polynucleotide and is also complementary to the sequence of the original polynucleotide. Illustratively, if the original polynucleotide sequence is ACTGAAC, then the reverse complementary sequence thereof is GTTCAGT.

The "internal ribosomal entry site" (IRES) in the present disclosure belongs to the translational control sequences and is usually located on the 5' end of the gene of interest, and enables the translation of RNA in a cap-independent manner. The transcribed IRES may directly bind to a ribosomal subunit, such that the initiation codon of an mRNA is appropriately oriented in a ribosome to perform translation. The IRES sequence is usually located in the 5'UTR of an mRNA (directly upstream of the initiation codon). IRES functionally replaces the need for various protein factors that interact with the translation mechanism of eukaryotes.

The "vector" in the present disclosure refers to a delivery vehicle for a polynucleotide. In some embodiments, the vector includes a polynucleotide sequence encoding a certain protein operatively inserted therein and enables the expression of this protein in a genetic engineering recombinant technique. The vector can be used to transform, transduce or transfect a host cell, and enable the genetic material element carried by the vector to be expressed in the host cell. The "vector" in the present disclosure may be any suitable vector, which includes chromosomal, non-chromosomal and synthetic nucleic acid vectors (including a group of suitable nucleic acid sequences which express the control elements). Illustratively, said vector may be a recombinant plasmid vector, a recombinant eukaryotic viral vector, a recombinant phage vector, a recombinant yeast minichromosome vector, a recombinant bacterial artificial chromosome vector, or a recombinant yeast plasmid vector.

Illustratively, the vector in the present disclosure may include the derivatives of SV40, bacterial plasmids, phage DNAs, baculovirus, yeast plasmid, vectors derived from a combination of a plasmid and a phage DNA, and vectors such as virus nuclear acids (RNA or DNA). In some embodiments, the vector is an adeno-associated virus (AAV) vector.

In a specific embodiment, the vectors related in the present disclosure are as shown in Figures 1 to 2, Figures 5 to 6 and Figures 8 to 9. In the schematic diagrams of the above-mentioned vectors, the meaning of CMV is human cytomegalovirus promoter (see e.g.,

PubMed PMID: 2985280), the meaning of TK pA is thymidine kinase polyadenylation signal (see e.g., PubMed PMID: 3018551), the meaning of SV40 is SV40 early promoter (see e.g., PubMed PMID: 6286831), the meaning of BSR is Blasticidin resistance gene: selection marker (see e.g., PubMed PMID: 7948022), the meaning of SV pA is SV40 polyadenylation signal (see e.g., PubMed PMID: 6113054), the meaning of pUC ori is the origin of the replication of pUC plasmid (see e.g., PubMed PMID: 2985470), the meaning of Amp is Ampicillin resistance gene (see e.g., PubMed PMID: 2985470), the meaning of EMCV IRES is the internal ribosomal entry site of encephalomyocarditis virus (see e.g., PubMed PMID: 8954121), the meaning of Zeocin is Zeocin resistance gene: selection marker (see e.g., PubMed PMID: 2450783), and the meaning of EF1αI is the first intron of the gene of human elongation factor 1 alpha (see e.g., PubMed PMID:2210382).

In some embodiments, the CMV promoter in the present disclosure can have a sequence that is at least or about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical to SEQ ID NO:16. As used herein, a promoter refers to a region of DNA that leads to initiation of transcription of a polynucleotide encoding a polypeptide. Promoters are located near the transcription start sites of the coding sequence, upstream on the DNA (towards the 5' region of the sense strand of the coding sequence). In some embodiments, other promotors can be used. In some embodiments, the promotor is a SV40 promoter, hCMV promoter, mCMV promoter, retinoschisin promoter, a rhodopsin promoter, a rhodopsin kinase promoter, a CRX promoter, or an interphotoreceptor retinoid binding protein (IRBP) promoter. Any promoter that allows tissue-specific expression of an encoded protein can also be used.

The “host cell” in the present disclosure refers to a cell having an heterologous polynucleotide and/or a vector introduced therein. Said host cell is a eukaryotic host cell or a prokaryotic host cell, wherein the eukaryotic host cell may be a mammalian host cell, an insect host cell, a plant host cell, a fungal host cell, a eukaryotic algae host cell, a nematode host cell, a protozoan host cell, and a fish host cell. Illustratively, the host cell in the present disclosure is a eukaryotic host cell, and said eukaryotic host cell is a mammalian host cell, wherein said mammalian host cell is selected from a Chinese hamster ovary cell (CHO cell), a COS cell, a Vero cell, a SP2/0 cell, a NS/O myeloma cell, a human embryonic kidney cell, an immature

hamster kidney cell, a HeLa cell, a human B cell, a cv-1/EBNA cell, an L cell, a 3T3 cell, a HEPG2 cell, a PerC6 cell, a 293 cell and an MDCK cell. Illustratively, the mammalian host cell in the present disclosure is a CHO cell.

The "protein expression system" in the present disclosure refers to a system comprising a host and a vector containing a heterologous gene, and the purpose of the expression of the heterologous gene in the host can be achieved by this system. The protein expression system generally comprises the following parts: (1) a host, i.e., an organism expressing proteins, which may be selected from bacteria, yeast, plant cells, animal cells, and the like; (2) a vector. The type of the vector matches with the host. According to the different hosts, the vectors are divided into prokaryotic (bacterial) expression vectors, yeast expression vectors, plant expression vectors, mammalian expression vectors, insect expression vectors, and the like. The vector contains a fragment of a heterologous gene. The heterologous gene can be expressed in the host via the mediation of the vector. In some embodiments, the expressed protein products are secreted. In some embodiments, the vectors are integrated into host cell DNA.

A key step in protein expression is the selection of recombinant host cells which have been successfully transfected with the vector comprising the heterologous gene coding the protein of interest. Most commonly a selection marker is included in the vector. The selection marker can be a gene or DNA sequence that allows separation of recombinant host cells containing the marker and those not containing it. The combination of a selection marker and a selection medium allows growth of recombinant host cells that have been transfected with the vector, while prohibiting the growth of host cells that have not been successfully transfected.

Antibiotic resistance genes are the most commonly used markers for recombinant host cell selection. An antibiotic resistance gene as a selection marker, in combination with a selection medium containing the antibiotic, can be used in order to achieve selection. Exemplary antibiotic selection markers include but are not limited to ampicillin resistance gene, chloramphenicol resistance gene, kanamycin resistance gene, tetracycline resistance gene, polymyxin B resistance gene, erythromycin resistance gene, carbenicillin resistance gene, streptomycin resistance gene, spectinomycin resistance gene, blasticidin resistance gene, neomycin resistance gene, puromycin resistance gene, zeocin resistance gene, and hygromycin B resistance gene. Accordingly, the selection antibiotics include but are not limited to

ampicillin, chloramphenicol, kanamycin, tetracycline, polymyxin B, erythromycin, carbenicillin, streptomycin, spectinomycin, blasticidin, neomycin, puromycin, zeocin, and hygromycin B. In some embodiments, the selection marker used in the present invention is blasticidin resistance gene. In some embodiments, the selection marker used in the present invention is zeocin resistance gene.

In some aspects, the disclosure provides methods that are designed for quickly evaluating a heteromultimer (e.g., antibody) expression. For example, for efficient expression of antibodies, the antibody heavy chain and the antibody light chain needs to be expressed in roughly 1:1 ratio. If the concentration for a selection antibiotic is too low, the amount of functional vectors in the cells can be too small. If the concentration for a selection antibiotic is too high, it may create a condition that is not favorable for culturing cells. Furthermore, the ratio of the two vectors needs to be properly adjusted. It has been determined, based on tests on many different conditions, the methods provided herein can express antibodies with a high efficiency, and can be used to reliably evaluate the heteromultimer expression in a reasonably short time. Furthermore, the methods provided herein can express antibodies with a high expression level.

In some embodiments, the methods involve transfecting the cell a pair of two vectors, one carrying a heterologous gene encoding a first polypeptide and the other carrying a heterologous gene encoding a second polypeptide. Two selection markers are used. One selection marker is blasticidin resistance gene, and the other selection marker is zeocin resistance gene. In some embodiment, blasticidin is present in the selection medium in an amount of 1-15 μg/mL and zeocin is present in an amount of 50-1500 μg/mL. Preferably, blasticidin is present in an amount of 2-12 μg/mL and zeocin is present in an amount of 100-1000 μg/mL. More preferably, blasticidin is present in an amount of 3-10 μg/mL and zeocin is present in an amount of 150-800 μg/mL. More preferably, blasticidin is present in an amount of 4-9 μg/mL and zeocin is present in an amount of 200-400 μg/mL. Most preferably, blasticidin is present in an amount of 9 μg/mL and zeocin is present in an amount of 400 μg/mL. Alternatively, blasticidin is present in an amount of 4 μg/mL and zeocin is present in an amount of 200 μg/mL. In some embodiments, the ratio of blasticidin concentration to Zeocin concentration is from 1:50 ~ 1:40 (e.g., about 9:400). In some embodiments, the minimum concertation of blasticidin in the medium is 5, 6, 7, 8, or 9 μg/mL. In some embodiments, the highest

concertation of blasticidin in the medium is 15 or 20 μg/mL.

In some embodiment, the selection medium further comprises serum, polysaccharide (e.g. glucose, and/or dextrose), sodium pyruvate, glutathione, ethanolamine, amino acid (e.g. glycine, alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, histidine, isoleucine, leucine, lysine, glutamine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and/or valine) or a salt thereof, vitamin (e.g. ascorbic acid phosphate, choline chloride, D-calcium pantothenate, folic acid, niacinamide, pyridoxine hydrochloride, riboflavin, thiamine hydrochloride, and/or i-inositol), inorganic salt (e.g. calcium chloride, ferric nitrate, magnesium sulfate, potassium chloride, sodium bicarbonate, sodium chloride, and/or sodium phosphate dibasic), protein (e.g. human transferrin and/or recombinant insulin), and/or trace element (e.g. ammonium metavanadate, cupric sulfate, manganous chloride, and/or sodium selenite).

In some embodiments, after about 18~30 hours (e.g., about 24 hours) of transfection, the cells are cultured in an appropriate cell culture medium containing blasticidin (e.g., 9 μg/mL) and Zeocin (e.g., 400 μg/mL). The cells are then passaged to a new medium containing blasticidin and Zeocin every 2 to 4 days. When the cell viability is recovered to 90% or more, the expression level of the heteromultimer can be evaluated by fed-batch cultures. In some embodiments, the fed-batch cultures can be any medium as described herein. In some embodiments, the fed-batch cultures contain blasticidin and Zeocin.

In some embodiments, the “protein expression system” in the present methods involve a pair of two vectors, one carrying a heterologous gene encoding an antibody heavy chain and the other carrying a heterologous gene encoding an antibody light chain. The selection marker in the two vectors might be different. In one embodiment, the selection marker in the first vector is blasticidin while the selection marker in the second vector is zeocin. The concentration of blasticidin and zeocin can be any concentrations as described herein. In some embodiments, the methods can also involve one vector comprising a heterologous gene encoding an antibody heavy chain and a heterologous gene encoding an antibody light chain.

The WXRE sequence can be inserted in the two vectors. They can be the same or different, and can have the forward or reverse directions. Table 1 lists the exemplary combinations of WXREs in the two vectors.

Table 1 Combinations of WXREs

#	WXRE in a first vector	WXRE in a second vector	#	WXRE in a first vector	WXRE in a second vector
1	SEQ ID NO 3	SEQ ID NO 3	54	SEQ ID NO 7	SEQ ID NO 17
2	SEQ ID NO 3	SEQ ID NO 4	55	SEQ ID NO 7	SEQ ID NO 18
3	SEQ ID NO 3	SEQ ID NO 5	56	SEQ ID NO 7	SEQ ID NO 19
4	SEQ ID NO 3	SEQ ID NO 6	57	SEQ ID NO 7	SEQ ID NO 20
5	SEQ ID NO 3	SEQ ID NO 7	58	SEQ ID NO 7	SEQ ID NO 21
6	SEQ ID NO 3	SEQ ID NO 8	59	SEQ ID NO 7	SEQ ID NO 22
7	SEQ ID NO 3	SEQ ID NO 9	60	SEQ ID NO 7	SEQ ID NO 23
8	SEQ ID NO 3	SEQ ID NO 17	61	SEQ ID NO 8	SEQ ID NO 8
9	SEQ ID NO 3	SEQ ID NO 18	62	SEQ ID NO 8	SEQ ID NO 9
10	SEQ ID NO 3	SEQ ID NO 19	63	SEQ ID NO 8	SEQ ID NO 17
11	SEQ ID NO 3	SEQ ID NO 20	64	SEQ ID NO 8	SEQ ID NO 18
12	SEQ ID NO 3	SEQ ID NO 21	65	SEQ ID NO 8	SEQ ID NO 19
13	SEQ ID NO 3	SEQ ID NO 22	66	SEQ ID NO 8	SEQ ID NO 20
14	SEQ ID NO 3	SEQ ID NO 23	67	SEQ ID NO 8	SEQ ID NO 21
15	SEQ ID NO 4	SEQ ID NO 4	68	SEQ ID NO 8	SEQ ID NO 22
16	SEQ ID NO 4	SEQ ID NO 5	69	SEQ ID NO 8	SEQ ID NO 23
17	SEQ ID NO 4	SEQ ID NO 6	70	SEQ ID NO 9	SEQ ID NO 9
18	SEQ ID NO 4	SEQ ID NO 7	71	SEQ ID NO 9	SEQ ID NO 17
19	SEQ ID NO 4	SEQ ID NO 8	72	SEQ ID NO 9	SEQ ID NO 18
20	SEQ ID NO 4	SEQ ID NO 9	73	SEQ ID NO 9	SEQ ID NO 19
21	SEQ ID NO 4	SEQ ID NO 17	74	SEQ ID NO 9	SEQ ID NO 20
22	SEQ ID NO 4	SEQ ID NO 18	75	SEQ ID NO 9	SEQ ID NO 21
23	SEQ ID NO 4	SEQ ID NO 19	76	SEQ ID NO 9	SEQ ID NO 22
24	SEQ ID NO 4	SEQ ID NO 20	77	SEQ ID NO 9	SEQ ID NO 23
25	SEQ ID NO 4	SEQ ID NO 21	78	SEQ ID NO 17	SEQ ID NO 17
26	SEQ ID NO 4	SEQ ID NO 22	79	SEQ ID NO 17	SEQ ID NO 18
27	SEQ ID NO 4	SEQ ID NO 23	80	SEQ ID NO 17	SEQ ID NO 19
28	SEQ ID NO 5	SEQ ID NO 5	81	SEQ ID NO 17	SEQ ID NO 20
29	SEQ ID NO 5	SEQ ID NO 6	82	SEQ ID NO 17	SEQ ID NO 21
30	SEQ ID NO 5	SEQ ID NO 7	83	SEQ ID NO 17	SEQ ID NO 22
31	SEQ ID NO 5	SEQ ID NO 8	84	SEQ ID NO 17	SEQ ID NO 23
32	SEQ ID NO 5	SEQ ID NO 9	85	SEQ ID NO 18	SEQ ID NO 18
33	SEQ ID NO 5	SEQ ID NO 17	86	SEQ ID NO 18	SEQ ID NO 19
34	SEQ ID NO 5	SEQ ID NO 18	87	SEQ ID NO 18	SEQ ID NO 20
35	SEQ ID NO 5	SEQ ID NO 19	88	SEQ ID NO 18	SEQ ID NO 21
36	SEQ ID NO 5	SEQ ID NO 20	89	SEQ ID NO 18	SEQ ID NO 22
37	SEQ ID NO 5	SEQ ID NO 21	90	SEQ ID NO 18	SEQ ID NO 23
38	SEQ ID NO 5	SEQ ID NO 22	91	SEQ ID NO 19	SEQ ID NO 19
39	SEQ ID NO 5	SEQ ID NO 23	92	SEQ ID NO 19	SEQ ID NO 20
40	SEQ ID NO 6	SEQ ID NO 6	93	SEQ ID NO 19	SEQ ID NO 21

#	WXRE in a first vector	WXRE in a second vector	#	WXRE in a first vector	WXRE in a second vector
41	SEQ ID NO 6	SEQ ID NO 7	94	SEQ ID NO 19	SEQ ID NO 22
42	SEQ ID NO 6	SEQ ID NO 8	95	SEQ ID NO 19	SEQ ID NO 23
43	SEQ ID NO 6	SEQ ID NO 9	96	SEQ ID NO 20	SEQ ID NO 20
44	SEQ ID NO 6	SEQ ID NO 17	97	SEQ ID NO 20	SEQ ID NO 21
45	SEQ ID NO 6	SEQ ID NO 18	98	SEQ ID NO 20	SEQ ID NO 22
46	SEQ ID NO 6	SEQ ID NO 19	99	SEQ ID NO 20	SEQ ID NO 23
47	SEQ ID NO 6	SEQ ID NO 20	100	SEQ ID NO 21	SEQ ID NO 21
48	SEQ ID NO 6	SEQ ID NO 21	101	SEQ ID NO 21	SEQ ID NO 22
49	SEQ ID NO 6	SEQ ID NO 22	102	SEQ ID NO 21	SEQ ID NO 23
50	SEQ ID NO 6	SEQ ID NO 23	103	SEQ ID NO 22	SEQ ID NO 22
51	SEQ ID NO 7	SEQ ID NO 7	104	SEQ ID NO 22	SEQ ID NO 23
52	SEQ ID NO 7	SEQ ID NO 8	105	SEQ ID NO 23	SEQ ID NO 23
53	SEQ ID NO 7	SEQ ID NO 9			

The “sequence identity” and the “percent identity” in the present disclosure refer to the percentage of the same (i.e., identical) nucleotides or amino acids between two or more polynucleotides or polypeptides. The sequence identity between two or more polynucleotides or polypeptides can be determined by the following method. The nucleotide sequences or the amino acid sequences of the polynucleotides or polypeptides are aligned and the number of positions containing the same nucleotide or amino acid residue in the aligned polynucleotides or polypeptides is scored and compared with the number of positions containing different nucleotides or amino acid residues in the aligned polynucleotides or polypeptides. The polynucleotides may differ at one position, for example, by containing different nucleotides (i.e., substitutions or mutations) or by the deletion of nucleotide(s) (i.e., the insertion of nucleotide(s) or the deletion of nucleotide(s) in one or two polynucleotides). The polypeptides may differ at one position, for example, by containing different amino acids (i.e., substitutions or mutations) or by the deletion of amino acid(s) (i.e., the insertion of amino acid(s) or the deletion of amino acid(s) in one or two polypeptides). The sequence identity can be calculated by dividing the number of positions containing the same nucleotide or amino acid residue by the total number of the amino acid residues in the polynucleotide or polypeptide. For example, the percent identity can be calculated by dividing the number of positions containing the same nucleotide or amino acid residue by the total number of the nucleotides or the amino acid residues in the polynucleotide or polypeptide and multiplying the result by 100.

The "abnormal expression of a target protein in the animal" in the present disclosure means that, as compared with the expression level of the target protein in the animal under normal condition, the expression level of the target protein in the animal to be tested shows an increase or a decline; or a protein that should not be expressed under normal condition is expressed, or a protein that should be expressed is not expressed. In a technical solution, said animal refers to a mammal. In another technical solution, said mammal is human.

The term "antibody" in the present disclosure refers to an immunoglobulin, a fragment thereof, or a derivative of them, and includes any polypeptide comprising an antigen-binding site, regardless of whether it is produced *in vitro* or *in vivo*. This term includes, but is not limited to, a polyclonal antibody, a monoclonal antibody, a monospecific antibody, a bispecific antibody, a trispecific antibody, a multispecific antibody, a non-specific antibody, a humanized antibody, a fully human antibody, a chimeric antibody, a single-domain antibody, a single-stranded antibody, a synthetic antibody, a recombinant antibody, a heterozygous antibody, a mutated antibody, and a grafted antibody. The term "antibody" also includes antibody fragments such as Fab, Fab', $F(ab')_2$, Fv, scFv, Fd, dAb, and other antibody fragments that retain the antigen-binding function. Typically, such fragments will include an antigen-binding fragment.

The "fusion protein" in the present disclosure refers to a molecule comprising two or more proteins or the fragments thereof which are linked by the covalent bond via their respective main chains of the peptides, and more preferably, the fusion protein is generated by the genetic expression of the polynucleotide molecules encoding these proteins. In a preferred embodiment, the fusion protein comprises an immunoglobulin domain. In a preferred embodiment, the fusion protein is an Fc-fusion protein.

Illustratively, the antibodies that may be used in the present disclosure include, but are not limited to, Adalimumab, Bezlotoxumab, Avelumab, Dupilumab, Durvalumab, Ocrelizumab, Brodalumab, Reslizumab, Olaratumab, Daratumumab, Elotuzumab, Necitumumab, Infliximab, Obiltoxaximab, Atezolizumab, Secukinumab, Mepolizumab, Nivolumab, Alirocumab, Evolocumab, Dinutuximab, Bevacizumab, Pembrolizumab, Ramucirumab, Vedolizumab, Siltuximab, Alemtuzumab, Trastuzumab, Pertuzumab, Obinutuzumab, Brentuximab, Raxibacumab, Belimumab, Ipilimumab, Denosumab, Ofatumumab, Besilesomab, Tocilizumab,

Canakinumab, Golimumab, Ustekinumab, Certolizumab, Catumaxomab, Eculizumab, Ranibizumab, Panitumumab, Natalizumab, Omalizumab, Cetuximab, Efalizumab, Ibritumomab, Fanolesomab, Tositumomab, Gemtuzumab, Palivizumab, Necitumumab, Basiliximab, Rituximab, Capromab, Satumomab, and Muromonab.

Illustratively, the fusion proteins that may be used in the present disclosure include, but are not limited to, Etanercept, Alefacept, Abatacept, Rilonacept, Romiplostim, Belatacept, and Aflibercept.

In one embodiment, the present disclosure relates to the stringency of hybridization conditions which is used to define the degree of complementarity of two polynucleotides. Alternatively, the above-mentioned polynucleotide may be selected from DNAs. The "stringency" used in the present disclosure refers to the temperature and the ionic strength condition during the hybridization and the presence or absence of certain organic solvents. The higher the stringency, the higher the degree of complementarity between the target nucleotide sequence and the marked polynucleotide sequence. "Stringent conditions" refers to the temperature and the ionic strength condition under which the nucleotide sequence merely having high-frequency complementary bases will hybridize. The term "hybridizes under high stringency or very high stringency conditions" used herein describes the conditions for hybridization and washing. The guidance for performing the hybridization reaction can be found in "Current Protocols in Molecular Biology", John Wiley and Sons, N.Y. (1989), 6.3.1-6.3.6. The specific hybridization conditions mentioned in the present disclosure are as follows: 1) high stringency hybridization conditions: 6X sodium chloride/sodium citrate (SSC) at about 45 °C, followed by one or more mashes in 0.2X SSC, 0.1% SDS at 65℃; 2) very high stringency hybridization conditions: 0.5M sodium phosphate, 7% SDS at 65℃, followed by one or more washes at 0.2X SSC, 1% SDS at 65 °C.

In a technical solution of the present disclosure, said WXRE sequence has a sequence identity of at least or about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% (including all the ranges and percentages between these values) with the sequence of any of SEQ ID NOs: 3-9. In some embodiments, the WXRE sequence has at least or about 1, 2, 3, 4, 5, 6, 7, 9, or 10 conservative mutation relative to a sequence of any of SEQ ID NOs: 3-9. In some embodiments, the WXRE

sequence differs from a sequence selected from SEQ ID NOs: 3-9 by at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. The WXRE sequence can have a forward direction or a reverse direction. As used herein, a sequence of interest has a forward direction when the sense strand (from 5' to 3') has a sequence that is identical to the sequence of interest. A sequence of interest has a reverse direction when the sense strand has a sequence that is reverse complementary to the sequence of interest. The sequences that are reverse complementary to SEQ ID NOs: 3-9 are set forth in SEQ ID NO: 17-23 respectively. In some embodiments, the disclosure provides a sequence that is at least or about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% (including all the ranges and percentages between these values) of a sequence selected from SEQ ID NO: 17-23.

In some embodiments, the WXRE sequence can increase the expression amount of an heterologous protein by about or at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80% (e.g., as compared to a control sequence without WXRE sequence).

In a technical solution of the present disclosure, said transcriptional regulatory element may also be selected from a nucleotide sequence comprising a WXRE sequence and other sequences that may increase the level of protein expression of a eukaryotic cell line. Illustratively, the transcriptional regulatory element in the present disclosure is a nucleotide sequence comprising a WXRE sequence and a sequence having sequence identity with the sequence as shown in SEQ ID:13 (the first intron of human EF1αI).

In a technical solution of the present disclosure, said sequence having sequence identity with the sequence as shown in SEQ ID:13 has a sequence identity of at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% (including all the ranges and percentages between these values) with the sequence as shown in SEQ ID:13 (the first intron of human EF1αI). In some embodiments, the sequence can increase the expression amount of an heterologous protein by about or at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80% (e.g., as compared to a control sequence without such sequence).

In some embodiments, the WXRE sequence, the promotor, and the polynucleotide encoding a polypeptide are operably linked together. In some embodiments, the WXRE

sequence, the promotor, the polynucleotide encoding a polypeptide, and one or more additional regulatory elements are operably linked together. In some embodiments, the one additional regulatory element is an intron of EF1αI (e.g., the first intron of human EF1αI). The WXRE sequence, the promotor, and the polynucleotide encoding a polypeptide that are operably linked together can have various orders. For example, the WXRE sequence can be located before the promotor (e.g., from 5' to 3' on the sense strand of the coding sequence) or after the polynucleotide encoding a polypeptide (e.g., from 5' to 3' on the sense strand of the coding sequence). In some embodiments, there are at least or about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1K, 2K, 3K, 4K or 5K nucleotides between the WXRE sequence and the promoter or between the WXRE sequence and the polynucleotide encoding a polypeptide. In some embodiments, there are no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1K, 2K, 3K, 4K or 5K nucleotides between the WXRE sequence and the promoter or between the WXRE sequence and the polynucleotide encoding a polypeptide. In some embodiments, the one or more additional regulatory elements are located between the promoter and the sequence encoding a polypeptide.

The present disclosure also provides methods of enhancing the expression of a recombinant protein or polypeptide. The methods involve inserting one or more vectors as described herein into cells (e.g., by transformation or transfection); culturing the cells in the medium; and recovering the recombinant protein or polypeptide so expressed.

The molecular biological methods adopted in the present disclosure may all be found in the corresponding methods described in the disclosed publications such as "Current Protocols in Molecular Biology" (published by Wiley), "Molecular Cloning: A Laboratory Manual" (published by Cold Spring Harbor Laboratory).

Examples

Other objects, features and advantages of the present disclosure will become apparent with the following detailed description. However, it should be understood that the detailed description and the specific examples (though representing the specific embodiments of the present disclosure) are provided for illustrative purposes only since various changes and modifications made within the spirit and scope of the present disclosure will become apparent

to those skilled in the art after reading this detailed description.

All reagents used in the examples can be purchased and obtained from commercial sources, unless otherwise emphasized.

Example 1: Construction of a vector library and construction of a stable pool which expresses green fluorescent protein

1.1 Preparation of a vector library containing a genomic fragment of Chinese hamster ovary cells

1.1.1 1 μg of the GFP-expressing vector (i.e., the vector as shown in Figure 1) was subjected to enzyme digestion with BamHI in the enzyme digestion kit (NEB) containing the restriction endonuclease BamHI so as to be linearized and stayed overnight at 37℃ (the composition and the contents of the reagents in the enzyme digestion reaction were as shown in Table 2), wherein BamHI could be replaced with any other endonucleases corresponding to a unique restriction site which existed in the upstream of a promoter corresponding to GFP.

The schematic diagram of the GFP-expressing vector was as shown in Figure 1.

Table 2 The composition and the contents of the reagents in the enzyme digestion reaction

Reaction components	Volume
NEB CutSmart Buffer (Cat# B7204S)	5 μL
BamHI	5 μL
GFP-expressing vector	1 μg
Ultrapure water	make up to a total volume of 50 μL

1.1.2 Approximate five million CHO host cells were harvested, a DNeasy Blood & Tissue Kit (QIAGEN) was used to extract the genomic DNA of the CHO host cells, and said genomic DNA was dissolved in 100 μL of elution buffer inside the above-mentioned kit.

1.1.3 Five micrograms of the genomic DNA was subjected to enzyme digestion with 100 units of restriction endonuclease BglII (NEB) or DpnII (NEB) (the composition and the contents of the reagents in the enzyme digestion reaction were as shown in Table 3). Other restriction endonucleases might also be used, as long as they matched with the cohesive ends of the endonucleases of the linearized vector in step 1.1.1.

Table 3 The composition and the contents of the reagents in the enzyme digestion reaction

Reaction components	Volume
NEB CutSmart Buffer (Cat# B7204S)	5 μL
BamHI	5 μL
CHO genomic DNA	1 μg
Ultrapure water	make up to a total volume of 50 μL

1.1.4 The linearized vector in 1.1.1 was treated with 2 units of calf intestinal alkaline phosphatase (NEB) at 37℃ for approximate 30 minutes. Other types of alkaline phosphatases could also be used.

1.1.5 The linearized GFP-expressing vector in 1.1.4 and the digested CHO genomic DNA in 1.1.3 were subjected to separation by agarose gel electrophoresis, respectively. The gel was cut to recover the fragments of the GFP-expressing vector and the 1-4kb fragments of the digested genome, DNA was extracted from the agarose gel after electrophoresis using a QIAquick Gel Extraction Kit (QIAGEN).

1.1.6 The fragments of the GFP-expressing vector and the fragments of the genome recovered in 1.1.5 were subjected to ligation using a DNA Ligation Kit (Takara, Cat# 6022) for 45 minutes at 16℃ (the composition and the contents of the reagents in the ligation reaction were as shown in Table 4).

Table 4 The composition and the contents of the reagents in the ligation reaction

Reaction components	Volume
the recovered CHO genomic DNA	4 μL
the recovered vector	6 μL
Solution I	20 μg
Ultrapure water	10 μL

1.1.7 Ten microliter of the ligation product obtained by 1.1.6 was taken, 100 μL of competent cells were added, put in the ice bath for 30 minutes, thermally stimulated at 42℃ for 1 minute, and then put on the ice for 1 minute. 500 μL of fresh LB medium free of antibiotic was added to each tube of cells, and the cells were subjected to a 45-minute-recovery at 37℃. The step of plating was skipped and 500 mL of medium containing 100mg/L of Ampicillin was added directly. The vector extraction was performed using a Plasmid Maxi Kit (QIAGEN). The extracted DNA was used as the vector library.

1.1.8 The vector library obtained in 1.1.7 was linearized using the restriction endonucleases in which the restriction sites were merely located in the prokaryotic region of the backbone of the vector (for example, PvuI (NEB)), and stayed overnight under the same reaction conditions as in 1.1.1 at 37℃. The DNA was recovered by the phenol-chloroform method and used for transfection the next day.

1.2 Construction of the stable pool which expresses green fluorescent protein

1.2.1 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in an Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 0.3 μg to 0.6 μg of the linearized vector library obtained by step 1.1.8 were mixed evenly, and the cells were resuspended with this mixed solution and transferred to an electroporation cuvette. The cells were subjected to transfection using a program corresponding to the respective host cells in a 4D-Nucleofector™ System electroporation instrument. The cells after electroporation were suspended with 5 mL of medium free of antibiotic and placed in a shaker at 37℃ for cultivation.

1.2.2 Twenty-four hours after the transfection, equal volume of selective medium containing antibiotic corresponding to the resistance gene in the vector was added in the cell culture (in this experiment, the antibiotic was 800 μg/mL of Zeocin).

1.2.3 The cells were counted every 2 to 4 days. Cell passage was performed according to the growth situation of the cells, and screening was performed by using the selective medium with antibiotic corresponding to the resistance gene in the vector (in this experiment, the antibiotic was 400 μg/mL of Zeocin). Clone screening was prepared when the cell viability recovered to 90% or more.

Example 2: Screening of a clone highly expressing green fluorescent protein

2.1 Single-cell sorting and expansion

2.1.1 The cells in the recovered stable pool in the step 1.2.3 of Example 1 with a higher GFP expression level (for example, the top 0.5% of the expression level) were sorted by a FACS AriaII flow cytometer into a 96-well plate for cultivation.

2.1.2 75% of the medium in the plate was changed every 2 to 4 days until the recovered cells were visible by naked eyes.

2.2 Screening for a clone highly expressing GFP

The cells recovered in 2.1.2 were successively transferred into a new 96-well plate respectively, altogether about 300 clones (all the cells in each well were derived from one cell and were referred to as a clone herein). The expression amount of GFP was determined by a FACS AriaII flow cytometer, and the clones whose detected intensities were among the top 10% were transferred to a 24-well plate for expansion.

Example 3: Screening, identification and verification of the transcriptional regulatory element

3.1 Identification of a candidate sequence of the transcriptional regulatory element

3.1.1 When the cells that were expanded to the 24-well plates in 2.2 substantially covered the bottom of the plate, and the DNeasy Blood & Tissue Kit (QIAGEN) was used to extract the genome of each clone.

3.1.2 A forward primer and a reverse primer were respectively designed in the vector (about 200bp away from the upstream and the downstream of the restriction site of BamHI), and the genomes extracted in 3.1.1 were successively subjected to PCR amplification, wherein the sequence of the forward primer of the PCR reaction was GCAAAAAAGGGAATAAGGGCGACACGG(SEQ ID NO:1) and the sequence of the reverse primer of the PCR reaction was CATAGCCCATATATGGAGTTCCGCGTTA(SEQ ID NO:2).

The reaction system of the above-mentioned PCR reaction was as shown in Table 5.

Table 5 The reaction system of the PCR reaction

Reaction components	Volume
5X Q5 Reaction Buffer	5 μL
10 mM dNTPs	0.5 μL
10 μM forward primer	1.25 μL
10 μM reverse primer	1.25 μL
genome	1 μL
Q5 DNA Polymerase (Cat# M0491S)	0.25 μl
Ultrapure water	15.75 μL

The reaction steps of the above-mentioned PCR reaction were as shown in Table 6.

Table 6 The reaction steps of the PCR reaction

Temperature	Time	Number of Cycles
98°C	1 min	1
98°C	30s	35
61°C	30s	
68°C	5 min	
68°C	10 min	1

3.1.3 PCR products were subjected to separation by agarose gel electrophoresis, the gel was cut to recover the specific band(s) of 1kb or more, and the QIAquick Gel Extraction Kit (QIAGEN) was used to extract DNA.

3.1.4 The recovered band(s) was sent for sequencing, and the sequence A~G of the candidate transcriptional regulatory elements were identified.

3.1.5 The sequence A~G of the transcriptional regulatory elements obtained by sequencing and identification were as follows, wherein

the sequence of the transcriptional regulatory element A was the sequence as shown in SEQ ID NO:3 (the reverse complementary sequence of SEQ ID NO: 3 is SEQ ID NO: 17),

the sequence of the transcriptional regulatory element B was the sequence as shown in SEQ ID NO:4 (the reverse complementary sequence of SEQ ID NO: 4 is SEQ ID NO: 18),

the sequence of the transcriptional regulatory element C was the sequence as shown in SEQ ID NO:5 (the reverse complementary sequence of SEQ ID NO: 5 is SEQ ID NO: 19),

the sequence of the transcriptional regulatory element D was the sequence as shown in SEQ ID NO:6 (the reverse complementary sequence of SEQ ID NO: 6 is SEQ ID NO: 20),

the sequence of the transcriptional regulatory element E was the sequence as shown in SEQ ID NO:7 (the reverse complementary sequence of SEQ ID NO: 7 is SEQ ID NO: 21),

the sequence of the transcriptional regulatory element F was the sequence as shown in SEQ ID NO:8 (the reverse complementary sequence of SEQ ID NO: 8 is SEQ ID NO: 22),

the sequence of the transcriptional regulatory element G was the sequence as shown in SEQ ID NO:9 (the reverse complementary sequence of SEQ ID NO: 9 is SEQ ID NO: 23).

3.2 Verification of the transcriptional regulatory element

3.2.1 The transcriptional regulatory element A~G obtained by sequencing and identification in 3.1.5 were respectively inserted into an upstream BamHI restriction site of the

corresponding promoter in a vector containing the GFP gene using an In-Fusion Cloning Kit (Takara). A vector with the transcriptional regulatory element inserted therein as shown in Figure 2 (wherein WXRE showed one of the transcriptional regulatory elements A~G) was obtained. The above-mentioned vector was linearized using the restriction endonucleases in which the restriction sites were merely located in the prokaryotic region of the backbone of the vector (for example, PvuI (NEB)), and stayed overnight at 37℃. DNA was recovered by phenol-chloroform and used for transfection the next day.

3.2.2 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in the Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 30 μg of a linearized vector containing the protein to be expressed (obtained by 3.2.1) were mixed evenly, and the cells were resuspended with this mixed solution and transferred to the electroporation cuvette. The cells were subjected to transfection using the program corresponding to the respective host cells in the 4D-Nucleofector™ System electroporation instrument. The cells after electroporation were resuspended with 5 mL of medium free of antibiotic and placed in a shaker at 37℃ for cultivation. Each sample contained one kind of transcriptional regulatory element or was a control without any transcriptional regulatory element.

3.2.3 Twenty-four hours after transfection, equal volume of selective medium containing antibiotic corresponding to the resistance gene in the vector was added in the cell culture. The cells were passaged using a medium containing antibiotic(s) every 2 to 4 days.

3.2.4 After the cell viability recovered to 90% or more, the influence of the transcriptional regulatory element A~G on the expression level of the proteins was evaluated by fed-batch cultures.

Example 4: Influence of the transcriptional regulatory elements on the expression level of a protein expression system used to express an heterologous protein

4.1.1 The transcriptional regulatory element A~G were respectively constructed into the upstream BamHI position of the promoter of a fusion protein (the above-mentioned fusion protein was the A chain of PD-L1, whose sequence was the sequence as shown in SEQ ID NO: 10) in both forward and reverse directions. A vector with the transcriptional regulatory element inserted therein as shown in Figure 2 (wherein WXRE showed one of the transcriptional

regulatory element A~G) was obtained, wherein the number after the element name (A~G) indicates the direction of the WXRE regulatory element. The number 1 after the name of the transcriptional regulatory element indicated the forward direction and the number 2 indicated the reverse direction. For example, transcriptional regulatory element A1 showed the forward (i.e., 5' to 3') sequence of the sequence as shown in SEQ ID NO: 3 in the sense strand of coding sequence. Transcriptional regulatory element A2 showed the reverse complementary sequence of the sequence as shown in SEQ ID NO: 17 in the sense strand of protein coding sequence.

The above-mentioned vectors were linearized using the restriction endonucleases in which the restriction sites were merely located in the prokaryotic region of the backbone of the vector (for example, PvuI (NEB)) and stayed overnight under a condition of 37℃. The DNA was recovered by phenol-chloroform and used for transfection the next day.

4.1.2 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in the Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 30 μg of the linearized vector containing the fusion protein (obtained by 4.1.1) were mixed evenly, and the cells were resuspended with this mixed solution and transferred to the electroporation cuvette. The cells were subjected to transfection using the program corresponding to the respective host cells in the 4D-Nucleofector™ System electroporation instrument. The cells after electroporation were resuspended with 5 mL of medium free of antibiotic and placed in a shaker at 37℃ for cultivation. Samples in each group only contained one transcriptional regulatory element in a certain direction (i.e., the forward direction or the reverse direction) and a sample which did not contain any transcriptional regulatory element was taken as a control.

4.1.3 Twenty-four hours after transfection, equal volume of medium containing 800 μg/mL of Zeocin was added into the transfected cells.

4.1.4 The cells were passaged using a medium containing 400 μg/mL of Zeocin every 2 to 4 days.

4.1.5 When the cell viability recovered to 90% or more, the expression level of the fusion protein PD-L1 was subjected to evaluation by fed-batch cultures.

4.1.6 Whether the sequence of the PD-L1 obtained by expression was identical to the sequence as shown in SEQ ID NO: 10 was verified.

4.2 Experimental results

As shown in Figure 3, as compared with the control group which did not have the transcriptional regulatory element, inserting the transcriptional regulatory element in the upstream of the promoter of the fusion protein could increase the expression amount of the target protein by about 10% to 25% (see A2, B1, B2, D2, E2, F2 and G1 in Figure 3). The promoting effect of the above-mentioned sequence on protein expression in a certain direction was superior to that in the other direction, which might be related to the directionality of the promoter.

As shown in Figure 4, corresponding to the expression amount, the forward direction or the reverse direction of the above-mentioned transcriptional regulatory element could enable an increase about 10% in specific productivity (see A2, B1, B2, D2, E2 and F2 in Figure 4).

Meanwhile, by verification, the sequence of the PD-L1 obtained by expression was identical to the sequence as shown in SEQ ID NO: 10.

Example 5: Influence of the transcriptional regulatory elements on the expression level of a protein expression system used to express Adalimumab

5.1.1 The reverse sequence of the transcriptional regulatory element A (A2), the forward sequence of the transcriptional regulatory element B (B1) and the forward sequence of the transcriptional regulatory element G (G1) were respectively constructed into the upstream of the promoter which was located at the upstream of the gene that could express Adalimumab by the In-Fusion Cloning kit of Takara (the specific conditions were as shown in Table 7). Vectors with the transcriptional regulatory element inserted therein as shown in Figure 5 and Figure 6 (wherein WXRE showed one of the transcriptional regulatory element A~G) were obtained respectively, wherein the “transcriptional regulatory element in the upstream of the heavy chain” was cloned into the vector as shown in Figure 5 and the “transcriptional regulatory element in the upstream of the light chain” was cloned into the vector as shown in Figure 6. Among them, the amino acid sequence of the heavy chain (HC) of Adalimumab in Figure 5 was as shown in SEQ ID NO:11 and the amino acid sequence of the light chain (LC) of Adalimumab in Figure 6 was as shown in SEQ ID NO:12.

The above-mentioned vector was linearized using the restriction endonucleases in which the restriction sites were merely located in the prokaryotic region of the backbone of the vector (for example, PvuI (NEB)), and stayed overnight at 37℃. The DNA was recovered by phenol-chloroform and used for transfection the next day.

Table 7 Corresponding transcriptional regulatory elements under different conditions

sample ID	transcriptional regulatory element in the upstream of the heavy chain	transcriptional regulatory element in the upstream of the light chain
1	B1	B1
2	B1	G1
3	G1	G1
4	B1	A2
5	G1	A2
6 (control)	N/A	N/A

5.1.2 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in the Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 30 μg of the linearized vector containing the sequence of Adalimumab (obtained by 5.1.1) were mixed evenly, and the cells were resuspended with this mixed solution and transferred to the electroporation cuvette. The cells were subjected to transfection using the program corresponding to the respective host cells in the 4D-Nucleofector™ System electroporation instrument. The cells after electroporation were resuspended with 5 mL of a medium free of antibiotic and placed in a shaker at 37℃ for cultivation. Samples in each group only contained one transcriptional regulatory element in a certain direction (i.e., the forward direction or the reverse direction), and a sample which did not contain any transcriptional regulatory element was taken as a control.

5.1.3 A method that was designed for quickly evaluating antibody expression was used. This method can ensure that the antibody heavy chain and light chain are roughly expressed in 1:1 ratio, and can reliably evaluate the antibody expression in a reasonably short time. Twenty-four hours after transfection, equal volume of medium containing 18 μg/mL of

blasticidin and 800 μg/mL of Zeocin was added into the transfected cells.

5.1.4 The cells were passaged using a medium containing 9 μg/mL of blasticidin and 400 μg/mL of Zeocin every 2 to 4 days.

5.1.5 When the cell viability recovered to 90% or more, the expression level of Adalimumab was subjected to evaluation by fed-batch cultures. Since both the heavy chain expression vector and the light chain expression vector of Adalimumab could be transfected into the same host cell, the heavy chain and the light chain of Adalimumab was capable of being expressed simultaneously. Since the heavy chain and the light chain mentioned above were capable of self-assembly in the host cells, a complete Adalimumab was obtained.

5.1.5 The biological activity of the obtained Adalimumab was determined.

5.2 Experimental results

As compared with the control group, in part of the forward sequences containing the transcriptional regulatory element B (see sample 1, 2 and 4), the expression level of Adalimumab had an increase of 10% to 20% (as shown in Figure 7).

By determining the biological activity of Adalimumab expressed by the present heterologous protein expression vector, it was found that its biological activity was identical to the biological activity of the known commercial Adalimumab.

Example 6: Influence of the combination of the transcriptional regulatory elements on the expression level of a protein expression system used to express Adalimumab

6.1.1 Two vectors were constructed respectively as illustrated in Figure 8 and Figure 9, wherein WXRE is the forward sequence of the transcriptional regulatory element B (B1), EF1αI is the sequence of the first intron of human EF1αI gene as shown in SEQ ID NO: 13, HC is the nucleic acid sequence encoding the heavy chain of Adalimumab, amino acid sequence of which is shown in SEQ ID NO: 11, and LC is the nucleic amino acid sequence encoding the light chain (LC) of Adalimumab, amino acid sequence of which is shown in SEQ ID NO: 12. After completion of the construction of the above-mentioned vectors, plasmid extraction was carried out using an MN kit, and the obtained plasmid was used for cell transfection.

6.1.2 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in the

Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 30 μg of the vector containing the sequence of Adalimumab (obtained by 6.1.1) were mixed evenly, and the cells were resuspended with this mixed solution and transferred to the electroporation cuvette. The cells were subjected to transfection using the program corresponding to the respective host cells in the 4D-Nucleofector™ System electroporation instrument. The cells after electroporation were resuspended with 5 mL of a medium free of antibiotic and placed in a shaker at 37℃ for cultivation. One group of samples comprised one transcriptional regulatory element, i.e., EF1αI intron, another group of samples comprised two elements, i.e., B1 and EF1αI intron, and a sample free of any transcriptional regulatory element was taken as a control.

6.1.3 24 hours after transfection, equal volume of medium containing 8 μg/mL of blasticidin and 400 μg/mL of Zeocin was added into the transfected cells.

6.1.4 The cells were passaged using a medium containing 4 μg/mL of blasticidin and 200 μg/mL of Zeocin every 2 to 4 days.

6.1.5 When the cell viability recovered to 90% or more, the expression level of Adalimumab was subjected to evaluation by fed-batch cultures. Since both the heavy chain expression vector and the light chain expression vector of Adalimumab could be transfected into the same host cell, the heavy chain and the light chain of Adalimumab was capable of being expressed simultaneously. Since the heavy chain and the light chain mentioned above were capable of self-assembly in the host cells, a complete Adalimumab was obtained.

6.1.6 The biological activity of the obtained Adalimumab was determined.

6.2 Experimental results

The experimental results of Example 6 were as shown in Table 8.

Table 8 Comparison of the relative expression levels of Adalimumab on Day 14 under different combination conditions of the transcriptional regulatory element

No.	Vector	Relative expression level (Day 14)	Change in expression amount
1	CMV (control)	1	
2	CMV-EF1αI	1.18	increase by 18%
3	WXRE-CMV-EF1αI	1.42	increase by 42%

As could be seen from Table 8, the combination comprising WXRE and EF1αI intron significantly increase the expression level of Adalimumab.

By determining the biological activity of Adalimumab expressed by the heterologous protein expression vector, it was found that its biological activity was identical to the biological activity of the known commercial Adalimumab. Thus, the proteins that were expressed by this vector were folded properly.

Example 7: Influence of the combination of the transcriptional regulatory elements on the expression level of a protein expression system used to express Pembrolizumab

7.1.1 Two vectors were constructed respectively as illustrated in Figure 8 and Figure 9, wherein WXRE is the forward sequence of the transcriptional regulatory element B (B1), EF1αI is the sequence of the first intron of human EF1αI gene as shown in SEQ ID NO: 13, HC is the nucleic acid sequence encoding the heavy chain of Pembrolizumab, amino acid sequence of which is shown in SEQ ID NO: 14, and LC is the nucleic amino acid sequence encoding the light chain (LC) of Pembrolizumab, amino acid sequence of which is shown in SEQ ID NO: 15. After completion of the construction of the above-mentioned vectors, plasmid extraction was carried out using an MN kit, and the obtained plasmid was used for cell transfection.

7.1.2 The following experimental steps were same as the experimental steps described in Example 6.1.2 to 6.1.6 of the present disclosure.

7.2 Experimental results

The experimental results of Example 7 were as shown in Table 9.

Table 9 Comparison of the relative expression levels of pembrolizumab on Day 14 under different combination conditions of the transcriptional regulatory element

No.	Vector	Relative expression level (Day 14)	Change in expression amount
1	CMV (control)	1	
2	WXRE-CMV-EF1αI	1.16	increase by 16%

As could be seen from Table 9, the combination comprising WXRE and EF1αI intron significantly increase the expression level of Pembrolizumab.

By determining the biological activity of Pembrolizumab expressed by the heterologous

protein expression vector, it was found that its biological activity was identical to the biological activity of the known commercial Pembrolizumab.

Example 8: Influence of different antibiotic concentrations on expression level of a protein expression system

8.1.1 Two sets of vectors were constructed respectively as illustrated in Figure 10 and Figure 11. One set is for Adalimumab and the other is for Pembrolizumab. After completion of vector construction, plasmid extraction was carried out using an MN kit, and the obtained plasmid was used for cell transfection.

8.1.2 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in the Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 30 μg of the vector containing the sequence of Adalimumab or Pembrolizumab were mixed evenly, and the cells were resuspended with this mixed solution and transferred to the electroporation cuvette. The cells were subjected to transfection using the program corresponding to the respective host cells in the 4D-Nucleofector™ System electroporation instrument. The cells after electroporation were resuspended with 5 mL of a medium free of antibiotic and placed in a shaker at 37℃ for cultivation.

8.1.3 24 hours after transfection, the cells were passaged using a selection medium containing different concentrations of Blasticidin and/or Zeocin every 2 to 4 days. The specific concentrations of Blasticidin and/or Zeocin are listed in Table 10.

8.1.4 When the cell viability recovered to 90% or more, the antibody expression level of was subjected to evaluation by fed-batch cultures. Since both the heavy chain expression vector and the light chain expression vector could be transfected into the same host cell, the antibody heavy chain and light chain were capable of being expressed simultaneously. Since the heavy chain and the light chain were capable of self-assembly in the host cells, a complete antibody was obtained.

8.1.5 The biological activity of the obtained antibodies was determined.

8.2 Experimental results

The experimental results of Example 8 were shown in Table 10.

Table 10 Comparison of expression levels of Adalimumab and Pembrolizumab on Day 14 under different antibiotic concentrations

		Blasticidin Concentration (μg/mL)	Zeocin Concentration (μg/mL)	Titer (g/L)
Adalimumab	1	9	0	0.66
	2	0	400	1.14
	3	9	400	1.35
	4	4	200	1.24
Pembrolizumab	1	9	0	0.69
	2	0	400	1.99
	3	9	400	2.27
	4	4	200	2.15

As could be seen from Table 10, the combination comprising Blasticidin and Zeocin attained a significantly increased expression level, as compared to the sole use of Blasticidin or Zeocin.

By determining the biological activity of Adalimumab or Pembrolizumab expressed by the heterologous protein expression vector, it was found that its biological activity was identical to the biological activity of the known commercial Adalimumab or Pembrolizumab. Thus, the proteins that were expressed by this vector were folded properly.

The above-mentioned examples of the present disclosure are examples provided for clearly illustrating the present disclosure only and are not limitations to the embodiments of the present disclosure. For those of ordinary skill in the art, other changes or variations in different forms may also be made on the basis of the above-mentioned specification. There is no need and no way to exhaust all of the embodiments herein. Any modifications, equivalent substitutions, improvements, and the like made within the spirit and principle of the present disclosure should all be included within the protection scope of the claims of the present disclosure.

CLAIMS

1. An isolated polynucleotide molecule comprising a nucleotide sequence selected from the group consisting of (i) to (iv):

(i) a sequence of any of SEQ ID NOs: 3-9;

(ii) a reverse complementary sequence of the sequence of any of SEQ ID NOs: 3-9;

(iii) a reverse complementary sequence of a sequence capable of hybridizing with the sequence of (i) or (ii) under a high stringency hybridization condition or a very high stringency hybridization condition; and

(iv) a sequence having at least 80% sequence identity, or at least 90% sequence identity, alternatively at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity, most preferably at least 99% sequence identity with the sequence of (i) or (ii).

2. A vector comprising a polynucleotide molecule according to claim 1.

3. The vector according to claim 2, wherein the vector is a recombinant expression vector.

4. The vector according to claim 2 or 3, further comprising a nucleotide sequence selected from the group consisting of (v) to (viii):

(v) a sequence of SEQ ID NO: 13;

(vi) a reverse complementary sequence of the sequence of SEQ ID NO: 13;

(vii) a reverse complementary sequence of a sequence capable of hybridizing with the sequence of (v) or (vi) under a high stringency hybridization condition or a very high stringency hybridization condition; and

(viii) a sequence having at least 80% sequence identity, or at least 90% sequence identity, alternatively at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity, most preferably at least 99% sequence identity with the sequence of (v) or (vi).

5. The vector according to any of claims 2-4, further comprising a CMV promoter; preferably, the CMV promoter comprises a nucleotide sequence of SEQ ID NO: 16, or the CMV promoter comprises a sequence having at least 80% sequence identity, or at least 90% sequence identity, alternatively at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity, most preferably at least 99%

sequence identity with the sequence of SEQ ID NO: 16.

6. The vector according to claim 5, comprising a combination of (1) a sequence of SEQ ID NO 4 or a sequence having 90% identity thereto; (2) a sequence of SEQ ID NO 13 or a sequence having 90% identity thereto; and (3) a sequence of SEQ ID NO 16 or a sequence having 90% identity thereto.

7. The vector according to any of claims 2-6, further comprising one or more genes encoding one or more proteins.

8. The vector according to claim 7, wherein the protein is selected from a group consisting of an antibody, a fusion protein, an enzyme, a soluble protein, a membrane protein, a structural protein, a ribosome protein, a zymogen, a cell surface receptor protein, a transcriptional regulatory protein, a translational regulatory protein, a chromatin protein, a hormone, a cell cycle regulatory protein, a G protein, a neuroactive peptide, an immunomodulatory protein, a blood component protein, an ion gate protein, a heat shock protein, dihydrofolate reductase, an antibiotic resistance protein, a functional fragment of any one of the proteins, an epitope fragment of any one of the proteins, and any combination thereof.

9. A recombinant host cell comprising the vector according to any of claims 2-8.

10. A method of preparing a recombinant host cell which stably expresses a protein, comprising a step of inserting into a host cell the vector according to any of claims 2-8.

11. The recombinant host cell according to claim 9 or the method according to claim 10, wherein the host cell is a Chinese hamster ovary (CHO) cell.

12. A method of preparing a protein, comprising a step of culturing the recombinant host cell according to claim 9 or 11 under conditions that allow production of the protein.

13. A kit comprising the recombinant host cell according to claim 9 or 11.

14. Use of the recombinant host cell according to claim 9 or 11 in the preparation of a reagent or a kit for detecting a disease due to the abnormality of protein expression.

15. Use of the recombinant host cell according to claim 9 or 11 in the preparation of a drug for treating or preventing a disease.

16. A method of screening a recombinant host cell which stably expresses a protein, comprising a step of screening a host cell using a selection medium comprising blasticidin and zeocin.

17. The method of claim 16, wherein blasticidin is present in an amount of 1-15 μg/mL and zeocin is present in an amount of 50-1500 μg/mL.

18. The method of claim 17, wherein blasticidin is present in an amount of 4-9 μg/mL and zeocin is present in an amount of 200-400 μg/mL.

19. The method of claim 18, wherein blasticidin is present in an amount of 9 μg/mL and zeocin is present in an amount of 400 μg/mL.

20. The method of claim 18, wherein blasticidin is present in an amount of 4 μg/mL and zeocin is present in an amount of 200 μg/mL.