

processes such as removal of virus particles by using filters or resins, as well as inactivation steps by using low pH or detergents, are implemented to prevent the serious safety issue of viral contamination of protein drug substances. Given the complexity of therapeutic proteins with respect to their large molecular size, post-translational modifications, and the variety of biological materials involved in their manufacturing process, the ability to enhance particular functional attributes while maintaining product safety and efficacy achieved through protein-engineering strategies is highly desirable.

While the integration of novel strategies and approaches to modify protein drug products is not a trivial matter, the potential therapeutic advantages have driven the increased use of such strategies during drug development. A number of protein-engineering platform technologies are currently in use to increase the circulating half-life, targeting, and functionality of novel therapeutic protein drugs as well as to increase production yield and product purity. For example, protein conjugation and derivatization approaches, including Fc-fusion, albumin-fusion, and PEGylation, are currently being used to extend a drug's circulating half-life.

The production of protein pharmaceutical (biologics) is expensive and time consuming. What is needed in the art are more efficient tools and processes for producing this important class of drugs.

SUMMARY

The present disclosure refers to nucleic acid constructs and their use to develop host cell lines for production of a protein of interest, and in particular to nucleic acid constructs which allow for improved selection to develop high-producing cell lines.

In some preferred examples, the present disclosure describes nucleic acid constructs for expression of a protein of interest comprising the following elements in operable association in 5' to 3' order: optionally, a first promoter sequence; a selectable marker sequence; a second promoter sequence; a nucleic acid sequence encoding a first protein of interest that is operably linked to the second promoter sequence; and a poly A signal sequence; the nucleic acid construct further comprising at least one insertion element at a position or positions selected from the group consisting of 5' to the optional first promoter or selectable marker sequence, 3' to the poly A signal sequence, between the optional first promoter and the poly A signal sequence, between the selectable marker and the second promoter sequence, and both 5' to the optional first promoter sequence or the selectable marker sequence and 3' to the poly A signal sequence.

In one aspect, the present disclosure refers to a nucleic acid construct for expression of a protein of interest comprising the following elements in operable association in 5' to 3' order: optionally, a first promoter sequence; a selectable marker sequence; a second promoter sequence; a nucleic acid sequence encoding a first protein of interest that is operably linked to the second promoter sequence; and a poly A signal sequence; the nucleic acid construct further comprising at least one insertion element at a position or positions selected from the group consisting of 5' to the optional first promoter or selectable marker sequence, 3' to the poly A signal sequence, between the optional first promoter and the poly A signal sequence, between the selectable marker and the second promoter sequence, and both 5' to the optional first promoter sequence or the selectable marker sequence and 3' to the poly A signal sequence. In some examples,

nucleic acid constructs comprise the first promoter sequence. In some preferred examples, the construct does not comprise a poly A signal sequence between the selectable marker and the second promoter. In some preferred examples, the selectable marker is adjacent to the second promoter. In some preferred examples, the second promoter is adjacent to the nucleic acid sequence encoding the first protein of interest. In some preferred examples, the nucleic acid construct comprises a non-coding region between the first promoter and the selectable marker. In some preferred examples, the non-coding region comprises multiple potential Kozak sequences and/or ATG translation start sites. In some preferred examples, the nucleic acid construct comprises an extending packaging region (EPR) between the first promoter and the selectable marker. In some preferred examples, the EPR comprises multiple potential Kozak sequences and/or ATG translation start sites.

In some preferred examples, the first promoter sequence is selected from the group consisting of SIN-LTR, SV40, *E. coli* lac, *E. coli* trp, phage lambda PL, phage lambda PR, T3, T7, cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, alpha-lactalbumin, human elongation factor 1 alpha (hEF1alpha), and mouse metallothionein-I promoter sequences. In some preferred examples, the first promoter sequence is not a retroviral LTR promoter.

In some preferred examples, the selectable marker sequence is an amplifiable selectable marker sequence selected from the group consisting of the Glutamine Synthase (GS) sequence and the Dihydrofolate Reductase (DHFR) sequence. In some preferred examples, the selectable marker sequence is an antibiotic resistance marker sequence selected from the group consisting of neomycin resistance gene (neo), hygromycin B phosphotransferase gene and puromycin N-acetyl transferase gene sequences.

In some preferred examples, the second promoter sequence is selected from the group consisting of SV40, *E. coli* lac, *E. coli* trp, phage lambda PL, phage lambda PR, T3, T7, cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, alpha-lactalbumin, human elongation factor 1 alpha (hEF1alpha), and mouse metallothionein-I promoter sequences.

In some preferred examples, the nucleic acid sequence encoding a protein of interest encodes a protein selected from the group consisting of heavy and light chain immunoglobulin sequences.

In some preferred examples, the insertion element is selected from the group consisting of a transposon insertion element, a recombinase insertion element, and a HDR

insertion element. In some preferred examples, the transposon insertion element is an inverted terminal repeat. In some preferred examples, the construct comprises two inverted terminal repeats positioned 5' to the first promoter and 3' to the poly A signal sequence. In some preferred examples, the recombinase insertion element is an attachment site (att). In some preferred examples, the attachment site (att) is attB. In some preferred embodiments, the HDR insertion element comprises an AAVS1 safe harbor locus sequence. In some preferred examples, the HDR insertion element is a nucleic acid sequence homologous to a target site in a chromosome. In some preferred examples, the nucleic acid sequence homologous to a target site in a chromosome is from about 30 to 1000 bases in length. In some preferred examples, the construct comprises two nucleic acid sequences homologous to a target site in a chromosome positioned 5' to the first promoter and 3' to the poly A signal sequence. In some preferred examples, the recombinase insertion element is a Flp Recombination Target (FRT) site. In some preferred examples, the recombinase insertion element is a LoxP sequence.

In some preferred examples, the constructs further comprise an RNA export element. In some preferred examples, the RNA export element is located 3' or 5' to the nucleic acid sequence encoding the protein of interest. In some preferred examples, the RNA export element is a pre-mRNA processing enhancer (PPE). In some preferred examples, the RNA export element is a posttranscriptional regulatory element (PRE). In some preferred examples, the PRE RNA export element is a Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE).

In some preferred examples, the constructs further comprise a signal peptide sequence operably linked to the first protein of interest. In some preferred examples, the signal peptide sequence is selected from the group consisting of tissue plasminogen activator, human growth hormone, lactoferrin, alpha-casein and alpha-lactalbumin signal peptide sequences.

In some preferred examples, the constructs further comprise a protein purification marker sequence. In some preferred examples, the protein purification marker sequence is a hexahistidine tag or a hemagglutinin (HA) tag.

In some preferred examples, the constructs further comprise an Internal Ribosome Entry Site (IRES) sequence and a second nucleic acid sequence encoding at least a second protein of interest (e.g., including third, fourth, fifth, etc. protein of interest) positioned 3' to the nucleic acid sequence encoding the first protein of interest. In some preferred

examples, the IRES sequence is selected from the group consisting of foot and mouth disease virus (FMDV), encephalomyocarditis virus and poliovirus IRES sequences.

In some preferred examples, the nucleic acid construct further comprises a third promoter operably linked to a second nucleic acid sequence encoding a second protein of interest positioned 3' to the nucleic acid sequence encoding the first protein of interest. In some preferred examples, the third promoter sequence is selected from the group consisting of SV40, *E. coli* lac, *E. coli* trp, phage lambda PL, phage lambda PR, T3, T7, cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, alpha-lactalbumin, human elongation factor 1 alpha (hEF1alpha), and mouse metallothionein-I promoter sequences. In some preferred examples, the constructs further comprise an RNA export element in operable association with the second nucleic acid sequence encoding a second protein of interest. In some preferred examples, the constructs further comprise a poly A signal sequence in operable association with the second nucleic acid sequence encoding a second protein of interest.

In some preferred examples, the first protein of interest is one of an antibody heavy and light chain and the second protein of interest is the other of an antibody heavy and light chain.

In some preferred examples, the nucleic acid construct further comprises an intron operably linked to a second nucleic acid sequence encoding a second protein of interest positioned 3' to the nucleic acid sequence encoding the first protein of interest. In some preferred examples, the constructs further comprise an RNA export element in operable association with the second nucleic acid sequence encoding a second protein of interest. In some preferred examples, the constructs further comprise a poly A signal sequence in operable association with the second nucleic acid sequence encoding a second protein of interest. In some preferred examples, the first protein of interest is one of an antibody heavy and light chain and the second protein of interest is the other of an antibody heavy and light chain.

In some preferred examples, the present disclosure describes a vector comprising a nucleic acid construct as described above. In some preferred examples, the vector is a plasmid.

In one aspect, the present disclosure refers to a host cell comprising the nucleic acid construct as disclosed herein. In some preferred examples, the present disclosure describes a host cell comprising a nucleic acid construct as described above or a vector as described above. In some preferred examples, the host cell is selected from the group consisting of Chinese Hamster Ovary (CHO) cells, HEK 293 cells, CAP cells, bovine mammary epithelial cells, monkey kidney

CV1 line transformed by SV40, baby hamster kidney cells, mouse sertoli cells, monkey kidney cells, African green monkey kidney cells, human cervical carcinoma cells, canine kidney cells, buffalo rat liver cells, human lung cells, human liver cells, mouse mammary tumor, TRI cells, MRC 5 cells, FS4 cells, rat fibroblasts, MDBK cells and human hepatoma line cells. In some preferred examples, the host cell is selected from the group consisting of a Chinese Hamster Ovary (CHO) cells, a HEK 293 cells and a CAP cells. In some preferred examples, the host cell line is a GS knockout cell line. In some preferred examples, the host cell line is a DHFR knockout cell line. In some preferred examples, the host cell comprises from about 1 to 1000 copies of the nucleic acid construct. In some preferred examples, the host cell comprises from about 10 to 200 copies of the nucleic acid construct. In some preferred examples, the host cell comprises from about 10 to 100 copies of the nucleic acid construct. In some preferred examples, the host cell comprises from about 20 to 100 copies of the nucleic acid construct. In some preferred examples, the host cell comprises from 50 to 500 copies of the nucleic acid construct. In some preferred examples, the host cell comprises from 50 to 250 copies of the nucleic acid construct.

In some preferred examples, the host cell further comprises at least a second nucleic acid construct that encodes and allows for expression of a second protein of interest. In some preferred examples, the second nucleic acid construct does not include a selectable marker. In some preferred examples, the second nucleic acid construct includes a selectable marker that is different from the selectable marker in the first nucleic acid construct. In some preferred examples, the first protein of interest in the first nucleic acid construct is one of an immunoglobulin heavy or light chain and the second protein in the second nucleic acid construct is the other of an immunoglobulin heavy or light chain. In some preferred examples, the first protein of interest is an immunoglobulin heavy chain and the second protein of interest is an immunoglobulin light chain. In some preferred examples, the host cell comprises from about 1 to 1000 copies of the second nucleic acid construct. In some preferred examples, the host cell comprises from about 10 to 200 copies of the second nucleic acid construct. In some preferred examples, the host cell comprises from about 10 to 100 copies of the second nucleic acid construct. In some preferred examples, the host cell comprises from about 20 to 100 copies of the second nucleic acid construct. In some preferred examples, the host cell comprises from 50 to 500 copies of the second nucleic acid construct. In some preferred examples, the host cell comprises from 50 to 250 copies of the second nucleic acid construct.

In some preferred examples, the present disclosure describes a host cell culture comprising a population of host cells as described above.

In another aspect, the present disclosure refers to a process for producing a protein of interest comprising culturing host cells as disclosed herein and purifying the protein of interest from the host cell culture. In some preferred examples, the present disclosure describes processes for producing a protein of interest comprising culturing host cells as described above under conditions such that the protein(s) of interest is expressed and purifying the protein(s) of interest from the host cell culture. In some preferred examples, the host cells grown in a medium comprising an inhibitor of the selectable marker. In some preferred examples, the selectable marker is GS and the inhibitor is phosphinothricin or methionine sulfoximine (Msx). In some preferred examples, the selectable marker is DHFR and the inhibitor is methotrexate.

In some preferred examples, the present disclosure describes a vector comprising the nucleic acid construct as described above. In some preferred examples, the vector is selected from the group consisting of a plasmid vector, a retroviral vector, a lentiviral vector, an AAV vector, and a transposon vector.

In one aspect, the present disclosure refers to a system comprising: a first nucleic acid construct as described herein; and a second nucleic acid construct encoding an enzyme. In some preferred examples, the constructs are provided on different vectors. In some preferred examples, the constructs are provided on the same vectors. In some preferred examples, the enzyme is selected from the group consisting of a transposase, an integrase, a recombinase, a nuclease and a nickase. In some preferred examples, the nuclease is a Cas nuclease. In some preferred examples, the nickase is a Cas nickase. In some preferred examples, the systems further comprise one or more RNA guide sequences. In some preferred examples, the enzyme facilitates insertion of the nucleic acid construct or a portion thereof into the genome of a host cell.

In some preferred examples, the systems further comprise at least a third nucleic acid construct as described above, the third nucleic acid construct encoding a protein of interest that is different from the protein of interest in the first nucleic acid construct. In some preferred examples, the third nucleic acid construct is provided in a separate vector. In some preferred examples, the third nucleic acid construct is provided in the same vector as the first and second nucleic acid constructs.

In some preferred examples, the present disclosure describes a system comprising at least first and second nucleic acid constructs as described above; wherein the first and second nucleic acid constructs each encode a different protein of interest. In some preferred examples, the first and second nucleic acid constructs are provided in separate vectors. In

some preferred examples, the first and second nucleic acid constructs are provided in the same vector. In some preferred examples, the systems further comprise a third nucleic acid construct encoding an enzyme. In some preferred examples, the enzyme is selected from the group consisting of a transposase, an integrase, a recombinase, a nuclease and a nickase. In some preferred examples, the nuclease is a Cas nuclease. In some preferred examples, the nickase is a Cas nickase. In some preferred examples, the systems further comprise one or more RNA guide sequences. In some preferred examples, the enzyme facilitates insertion of the nucleic acid construct or a portion thereof into the genome of a host cell. In some preferred examples, the third nucleic acid construct is provided in a separate vector. In some preferred examples, the third nucleic acid construct is provided in the same vector as the first and second nucleic acid constructs.

In some preferred examples, the present disclosure describes processes for producing a protein of interest comprising: introducing a nucleic acid construct, vector, or a system as described above into a host cell under conditions such that the nucleic acid construct is incorporated into the genome of the host cell; developing a host cell line that expresses the protein of interest; culturing host cells from the host cell line under conditions such that the protein of interest is produced by the host cells; and purifying the protein of interest from the host cell culture.

In some preferred examples, the host cell is selected from the group consisting of Chinese Hamster Ovary (CHO) cells, HEK 293 cells, CAP cells, bovine mammary epithelial cells, monkey kidney CV1 line transformed by SV40, baby hamster kidney cells, mouse sertoli cells, monkey kidney cells, African green monkey kidney cells, human cervical carcinoma cells, canine kidney cells, buffalo rat liver cells, human lung cells, human liver cells, mouse mammary tumor, TRI cells, MRC 5 cells, FS4 cells, rat fibroblasts, MDBK cells and human hepatoma line cells. In some preferred examples, the host cell is selected from the group consisting of a Chinese Hamster Ovary (CHO) cells, a HEK 293 cells and a CAP cells. In some preferred examples, the host cell line is a GS knockout cell line. In some preferred examples, the host cell line is a DHFR knockout cell line. In some preferred examples, the host cells are grown in a medium comprising an inhibitor of the selectable marker. In some preferred examples, the selectable marker is GS and the inhibitor is phosphinothricin or methionine sulfoximine (Msx). In some preferred examples, the selectable marker is DHFR and the inhibitor is methotrexate. In some preferred examples, culturing host cells from the host cell line under conditions such that the protein of interest is produced by the host cells further comprises culturing in a system selected from the group

CLAIMS

What is claimed is:

- 5 1. A nucleic acid construct for expression of a protein of interest comprising the following elements in operable association in 5' to 3' order:
 - optionally, a first promoter sequence;
 - a selectable marker sequence;
 - a second promoter sequence;
- 10 a nucleic acid sequence encoding a first protein of interest that is operably linked to the second promoter sequence; and
a poly A signal sequence;
the nucleic acid construct further comprising at least one insertion element at a position or positions selected from the group consisting of 5' to the optional first promoter or selectable
15 marker sequence, 3' to the poly A signal sequence, between the optional first promoter and the poly A signal sequence, between the selectable marker and the second promoter sequence, and both 5' to the optional first promoter sequence or the selectable marker sequence and 3' to the poly A signal sequence.
- 20 2. The nucleic acid construct of claim 1, wherein the nucleic acid construct does not comprise a poly A signal sequence between the selectable marker and the second promoter.
3. The nucleic acid construct of claim 1, wherein the selectable marker is adjacent to the second promoter.
- 25 4. The nucleic acid construct of claim 1, wherein the second promoter is adjacent to the nucleic acid sequence encoding the first protein of interest.
5. The nucleic acid construct of claim 1, wherein the nucleic acid construct comprises an
30 extending packaging region between the first promoter and the selectable marker.
6. The nucleic acid construct of claim 5, wherein the EPR comprises multiple potential Kozak sequences and/or ATG translation start sites.

7. The nucleic acid construct of claim 1, wherein the first promoter sequence is a weak promoter sequence.
8. The nucleic acid construct of claim 1, wherein the first promoter sequence is selected from the group consisting of SIN-LTR, SV40, *E. coli* lac, *E. coli* trp, phage lambda PL, phage lambda PR, T3, T7, cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, alpha-lactalbumin, human elongation factor 1 alpha (hEF1alpha), and mouse metallothionein-I promoter sequences.
9. The nucleic acid construct of claim 1, wherein the selectable marker sequence is an amplifiable selectable marker sequence selected from the group consisting of the Glutamine Synthase (GS) sequence and the Dihydrofolate Reductase (DHFR) sequence.
10. The nucleic acid construct of claim 1, wherein the selectable marker sequence is an antibiotic resistance marker sequence selected from the group consisting of neomycin resistance gene (neo), hygromycin B phosphotransferase gene and puromycin N-acetyl transferase gene sequences.
11. The nucleic acid construct of claim 1, wherein the second promoter sequence is selected from the group consisting of SV40, *E. coli* lac, *E. coli* trp, phage lambda PL, phage lambda PR, T3, T7, cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, alpha-lactalbumin, human elongation factor 1 alpha (hEF1alpha), and mouse metallothionein-I promoter sequences.
12. The nucleic acid construct of claim 1, wherein the nucleic acid sequence encoding a protein of interest encodes a protein selected from the group consisting of heavy and light chain immunoglobulin sequences.
13. The nucleic acid construct of claim 1 wherein the insertion element is selected from the group consisting of a transposon insertion element, a recombinase insertion element, and a HDR insertion element.
14. The nucleic acid construct of claim 13, wherein the transposon insertion element is an inverted terminal repeat.

15. The nucleic acid construct of claim 14, wherein the construct comprises two inverted terminal repeats positioned 5' to the first promoter and 3' to the poly A signal sequence.
- 5 16. The nucleic acid construct of claim 13, wherein the recombinase insertion element is an attachment site (att).
17. The nucleic acid construct of claim 16, wherein the attachment site (att) is attB.
- 10 18. A host cell comprising the nucleic acid construct of claim 1.
19. The host cell of claim 18, wherein the host cell comprises from about 1 to 1000 copies of the nucleic acid construct.
- 15 20. A process for producing a protein of interest comprising culturing host cells according to claim 18 and purifying the protein of interest from the host cell culture.
21. A system comprising:
a first nucleic acid construct according to claim 1; and
20 a second nucleic acid construct encoding an enzyme.
22. The system of claim 21, wherein the enzyme is selected from the group consisting of a transposase, an integrase, a recombinase, a nuclease and a nickase.