Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2685326 A1 2008/11/06

(21) 2 685 326

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2008/04/25
(87) Date publication PCT/PCT Publication Date: 2008/11/06
(85) Entrée phase nationale/National Entry: 2009/10/26
(86) N° demande PCT/PCT Application No.: US 2008/061483
(87) N° publication PCT/PCT Publication No.: 2008/134461
(30) Priorité/Priority: 2007/04/27 (US60/914,361)

(51) Cl.Int./Int.Cl. *C12N 15/78* (2006.01),
C12Q 1/68 (2006.01)

(71) Demandeur/Applicant:
DOW GLOBAL TECHNOLOGIES INC., US

(72) Inventeurs/Inventors:
RAMSEIER, THOMAS M., US;
COLEMAN, RUSSELL J., US;
SCHNEIDER, JANE C., US;
HERSHBERGER, CHARLES D., US

(74) Agent: SMART & BIGGAR

(54) Titre : PROCEDE POUR RAPIDEMENT CRIBLER DES HOTES MICROBIENS ET IDENTIFIER CERTAINES SOUCHES AYANT UN RENDEMENT ET/OU UNE QUALITE D'EXPRESSION DES PROTEINES HETEROLOGUES AMELIORES

(54) Title: METHOD FOR RAPIDLY SCREENING MICROBIAL HOSTS TO IDENTIFY CERTAIN STRAINS WITH IMPROVED YIELD AND/OR QUALITY IN THE EXPRESSION OF HETEROLOGOUS PROTEINS

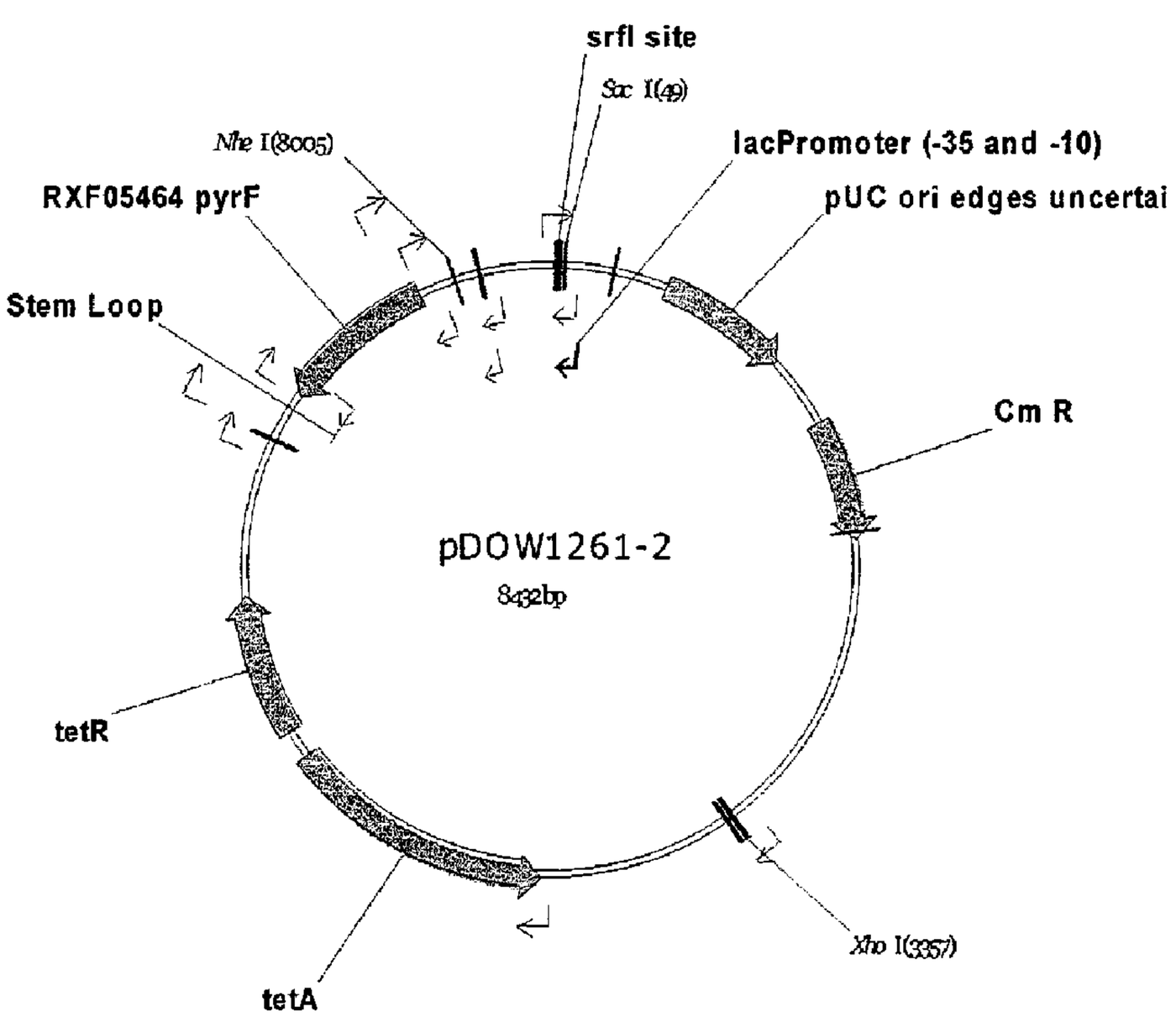

FIG. 1A

(57) Abrégé/Abstract:
The present invention provides an array for rapidly identifying a host cell population capable of producing heterologous protein with improved yield and/or quality. The array comprises one or more host cell populations that have been genetically modified to

(21) 2 685 326

(13) A1

(57) Abrégé(suite)/Abstract(continued):
increase the expression of one or more target genes involved in protein production, decrease the expression of one or more target genes involved in protein degradation, or both. One or more of the strains in the array may express the heterologous protein of interest in a periplasm compartment, or may secrete the heterologous protein extracellularly through an outer cell wall. The strain arrays are useful for screening for improved expression of any protein of interest, including therapeutic proteins, hormones, a growth factors, extracellular receptors or ligands, proteases, kinases, blood proteins, chemokines, cytokines, antibodies and the like.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 November 2008 (06.11.2008)

PCT

(10) International Publication Number
WO 2008/134461 A3

(51) International Patent Classification:
C12N 15/78 (2006.01) *C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/US2008/061483

(22) International Filing Date: 25 April 2008 (25.04.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/914,361 27 April 2007 (27.04.2007) US

(71) Applicant *(for all designated States except US)*: **DOW GLOBAL TECHNOLOGIES, INC.** [US/US]; 2040 Dow Center, Midland, MI 48674 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **RAMSEIER, Thomas, M.** [US/US]; 13033 Pomard Way, Poway, CA 92064 (US). **COLEMAN, Russell, J.** [US/US]; 1050 Island Avenue, #519, San Diego, CA 92101 (US). **SCHNEIDER, Jane, C.** [US/US]; 2621 32nd Street, San Diego, CA 92104 (US). **HERSHBERGER, Charles, D.** [US/US]; 15042 Amso Street, Poway, CA 92064 (US).

(74) Agents: **DAVENPORT, Destiny, M.** et al.; Alston & Bird LLP, Bank Of America Plaza, 101 South Tryon Street, Suite 4000, Charlotte, NC 28280-4000 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

[Continued on next page]

(54) Title: METHOD FOR RAPIDLY SCREENING MICROBIAL HOSTS TO IDENTIFY CERTAIN STRAINS WITH IMPROVED YIELD AND/OR QUALITY IN THE EXPRESSION OF HETEROLOGOUS PROTEINS

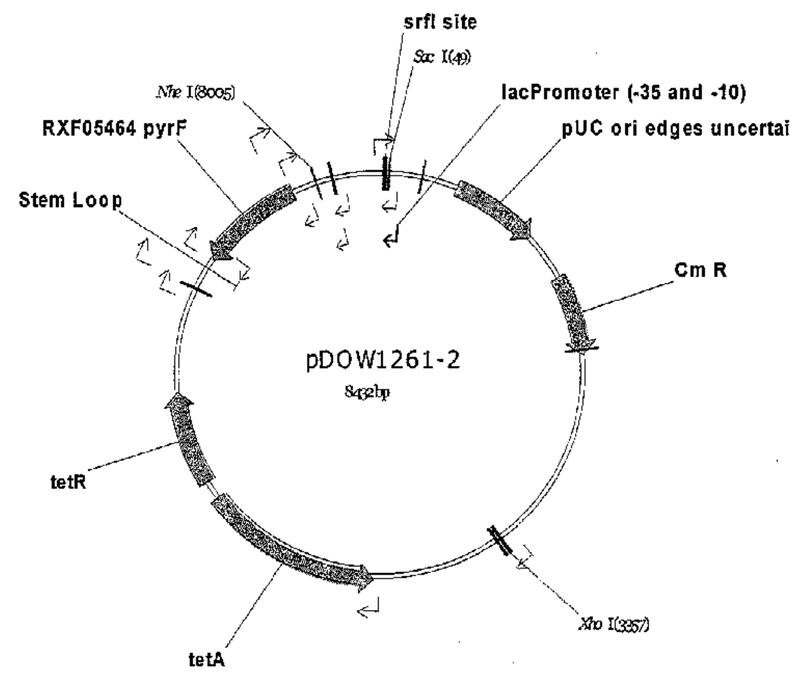

FIG. 1A

(57) Abstract: The present invention provides an array for rapidly identifying a host cell population capable of producing heterologous protein with improved yield and/or quality. The array comprises one or more host cell populations that have been genetically modified to increase the expression of one or more target genes involved in protein production, decrease the expression of one or more target genes involved in protein degradation, or both. One or more of the strains in the array may express the heterologous protein of interest in a periplasm compartment, or may secrete the heterologous protein extracellularly through an outer cell wall. The strain arrays are useful for screening for improved expression of any protein of interest, including therapeutic proteins, hormones, a growth factors, extracellular receptors or ligands, proteases, kinases, blood proteins, chemokines, cytokines, antibodies and the like.

WO 2008/134461 A3

WO 2008/134461 A3

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report:
24 December 2008

METHOD FOR RAPIDLY SCREENING MICROBIAL HOSTS TO IDENTIFY CERTAIN STRAINS WITH IMPROVED YIELD AND/OR QUALITY IN THE EXPRESSION OF HETEROLOGOUS PROTEINS

FIELD OF THE INVENTION

This invention is in the field of protein production, particularly to identifying optimal host cells for large-scale production of properly processed heterologous proteins.

BACKGROUND OF THE INVENTION

More than 150 recombinantly produced proteins and polypeptides have been approved by the U.S. Food and Drug Administration (FDA) for use as biotechnology drugs and vaccines, with another 370 in clinical trials. Unlike small molecule therapeutics that are produced through chemical synthesis, proteins and polypeptides are most efficiently produced in living cells. However, current methods of production of recombinant proteins in bacteria often produce improperly folded, aggregated or inactive proteins, and many types of proteins require secondary modifications that are inefficiently achieved using known methods.

Numerous attempts have been developed to increase production of properly folded proteins in recombinant systems. For example, investigators have changed fermentation conditions (Schein (1989) Bio/Technology, 7:1141-1149), varied promoter strength, or used overexpressed chaperone proteins (Hockney (1994) Trends Biotechnol. 12:456-463), which can help prevent the formation of inclusion bodies.

Strategies have been developed to excrete proteins from the cell into the supernatant. For example, U.S. Pat. No. 5,348,867; U.S. Pat. No. 6,329,172; PCT Publication No. WO 96/17943; PCT Publication No. WO 02/40696; and U.S. Application Publication 2003/0013150. Other strategies for increased expression are directed to targeting the protein to the periplasm. Some investigations focus on non-

Sec type secretion (see for e.g. PCT Publication No. WO 03/079007; U.S. Publication No. 2003/0180937; U.S. Publication No. 2003/0064435; and, PCT Publication No. WO 00/59537). However, the majority of research has focused on the secretion of exogenous proteins with a Sec-type secretion system.

A number of secretion signals have been described for use in expressing recombinant polypeptides or proteins. See, for example, U.S. Pat. No. 5,914,254; U.S. Pat. No. 4,963,495; European Patent No. 0 177 343; U.S. Pat. No. 5,082,783; PCT Publication No. WO 89/10971; U.S. Pat. No. 6,156,552; U.S. Pat. Nos. 6,495,357; 6,509,181; 6,524,827; 6,528,298; 6,558,939; 6,608,018; 6,617,143; U.S. Pat. Nos. 5,595,898; 5,698,435; and 6,204,023; U.S. Pat. No. 6,258,560; PCT Publication Nos. WO 01/21662, WO 02/068660 and U.S. Application Publication 2003/0044906; U.S. Pat. No. 5,641,671; and European Patent No. EP 0 121 352.

Heterologous protein production often leads to the formation of insoluble or improperly folded proteins, which are difficult to recover and may be inactive. Furthermore, the presence of specific host cell proteases may degrade the protein of interest and thus reduce the final yield. There is no single factor that will improve the production of all heterologous proteins. As a result, there is a need in the art for identifying improved large-scale expression systems capable of secreting and properly processing recombinant polypeptides to produce transgenic proteins in properly processed form.

SUMMARY OF THE INVENTION

The present invention provides compositions and methods for rapidly identifying a host cell population capable of producing at least one heterologous polypeptide according to a desired specification with improved yield and/or quality. The compositions comprise two or more host cell populations that have been genetically modified to increase the expression of one or more target genes involved in protein production, decrease the expression of one or more target genes involved in protein degradation, express a heterologous gene that affects the protein product, or a combination. The ability to express a polypeptide of interest in a variety of modified host cells provides a rapid and efficient means for determining an optimal host cell for the polypeptide of interest. The desired specification will vary depending on the polypeptide of interest, but includes yield, quality, activity, and the like.

It is recognized that the host cell populations may be modified to express many combinations of nucleic acid sequences that affect the expression levels of endogenous sequences and/or exogenous sequences that facilitate the expression of the polypeptide of interest. In one embodiment, two or more of the host cell populations has been genetically modified to increase the expression of one or more target genes involved in one or more of the proper expression, processing, and/or translocation of a heterologous protein of interest. In another embodiment, the target gene is a protein folding modulator. In another embodiment, the protein folding modulator is selected from the list in Table 1.

In another embodiment, one or more of the host cell populations has been genetically modified to decrease the expression of one or more target genes involved in proteolytic degradation. In another embodiment, the target gene is a protease. In another embodiment, the protease is selected from the list in Table 2.

In one embodiment, nucleotide sequences encoding the proteins of interest are operably linked to a *P. fluorescens* Sec system secretion signal as described herein. One or more of the strains in the array may express the heterologous protein of interest in a periplasm compartment. In certain embodiments, one or more strains may also secrete the heterologous protein extracellularly through an outer cell wall.

Host cells include eukaryotic cells, including yeast cells, insect cells, mammalian cells, plant cells, etc., and prokaryotic cells, including bacterial cells such as *P. fluorescens*, *E. coli*, and the like.

As indicated, the library of host cell populations can be rapidly screened to identify certain strain(s) with improved yield and/or quality of heterologously expressed protein. The strain arrays are useful for screening for improved expression of any protein of interest, including therapeutic proteins, hormones, a growth factors, extracellular receptors or ligands, proteases, kinases, blood proteins, chemokines, cytokines, antibodies and the like.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A depicts plasmid pDOW1261-2 used for engineering genomic deletion in *P. fluorescens*. Figure 1B is a schematic drawing of the constructions of a gene X deletion.

Figure 2 is a Western blot analysis of soluble cells fractions prepared at 0 and 24 hours post-induction (I0 and I24, respectively) in Δ*prc1, ΔdegP2, ΔLa2* and the *grpEdnaKJ* co-expression strains (Example 6). The top arrows point to the fully assembled monoclonal antibody in the co-expressed strains but not in the control (DC440). r = recombinant; n-r = nonrecombinant.

DETAILED DESCRIPTION

The present inventions now will be described more fully hereinafter with reference to the accompanying drawings, in which some, but not all embodiments of the invention are shown. Indeed, these inventions may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements.

Many modifications and other embodiments of the inventions set forth herein will come to mind to one skilled in the art to which these inventions pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings.

Therefore, it is to be understood that the inventions are not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the invention. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

Overview

Compositions and methods for identifying an optimal host strain, e.g, a *Pseudomonas fluorescens* host strain, for producing high levels of properly processed heterologous polypeptides in a host cell are provided. In particular, a library (or "array") of host strains is provided, wherein each strain (or "population of host cells") in the library has been genetically modified to modulate the expression of one or more target genes in the host cell. An "optimal host strain" can be identified or selected

based on the quantity, quality, and/or location of the expressed protein of interest compared to other populations of phenotypically distinct host cells in the array. Thus, an optimal host strain is the strain that produces the polypeptide of interest according to a desired specification. While the desired specification will vary depending on the polypeptide being produced, the specification includes the quality and/or quantity of protein, whether the protein is sequestered or secreted, protein folding, and the like.

"Heterologous," "heterologously expressed," or "recombinant" generally refers to a gene or protein that is not endogenous to the host cell or is not endogenous to the location in the native genome in which it is present, and has been added to the cell by infection, transfection, microinjection, electroporation, microprojection, or the like.

One or more of the host cell populations in the array is modified to modulate the expression of one or more target genes in the host cell. By "target gene" is intended a gene that affects heterologous protein production in a host cell. Target genes that affect heterologous protein production include genes encoding proteins that modulate expression, activity, solubility, translocation, proteolytic degradation and/or cleavage of the heterologous protein. For example, a target gene may encode at least one of a host cell protease, a protein folding modulator, a transcription factor, a translation factor, a secretion modulator, or any other protein involved in the proper transcription, translation, processing, and/or translocation of a heterologous protein of interest. A "target protein" refers to the protein or polypeptide resulting from expression of the target gene. Expression and/or activity of a target gene or genes is increased or decreased, depending on the function of the target gene or protein. For example, expression of one or more host cell proteases may be decreased, whereas expression of one or more protein folding modulators may be increased.

The arrays described herein are useful for rapidly identifying an optimal host cell for production of a heterologous protein or peptide of interest. Heterologous protein production often leads to the formation of insoluble or improperly folded proteins, which are difficult to recover and may be inactive. Furthermore, the presence of specific host cell proteases may degrade the protein of interest and thus reduce the final yield. There is no single host cell population that will optimally produce all polypeptides or proteins of interest. Thus, using the compositions and methods of the invention, an optimal host cell can be rapidly and efficiently identified from the library of modified cell populations. The optimal host strain can then be

used to produce sufficient amounts of the protein of interest or for commercial production. Likewise, a host strain can be modified for expression of the protein of interest based on the optimal host strain.

In one embodiment, the method includes obtaining an array comprising at least a first and a second population of *P. fluorescens* cells, wherein each population is selected from the group consisting of (i) a population of *P. fluorescens* cells that has been genetically modified to reduce the expression of at least target gene involved in protein degradation; (ii) a population of *P. fluorescens* cells that has been genetically modified to increase the expression of at least one target gene involved in protein production; and, (iii) a population of *P. fluorescens* cells that has been genetically modified to reduce the expression of at least one target gene involved in protein degradation and to increase the expression of at least target gene involved in protein production; introducing into at least one cell of each population an expression construct comprising at least one gene encoding at least one heterologous protein of interest; maintaining said cells under conditions sufficient for the expression of said protein of interest in at least one population of cells; and selecting the optimal population of cells in which the heterologous protein of interest is produced; wherein each population in the array is non-identical and wherein each population is physically separate one from another; wherein the heterologous protein of interest exhibits one or more of improved expression, improved activity, improved solubility, improved translocation, or reduced proteolytic degradation or cleavage in the optimal population of cells compared to other populations in the array.

The array may further comprise a population of host cells (e.g., *P. fluorescens* host cells) that has not been genetically modified to alter the expression of a host cell protease or a protein folding modulator. This population may be a wild-type strain, or may be a strain that has been genetically modified to alter the expression of or or more genes not involved in protein production, processing, or translocation (e.g., may be genetically modified to express, for example, a selectable marker gene).

In one embodiment, each population of *P. fluorescens* host cells is phenotypically distinct (i.e., "non-identical") one from another. By "phenotypically distinct" is intended that each population produces a measurably different amount of one or more target proteins. In this embodiment, each strain has been genetically modified to alter the expression of one or more different target genes. Where the expression of more than one target gene is modulated in a population of host cells,

then the combination of target genes is phenotypically distinct from other populations in the library. An array comprising a plurality of phenotypically distinct populations of host cells according to the present invention is one that provides a diverse population from which to select one or more strains useful for producing a heterologous protein or peptide of interest. It will be understood by one of skill in the art that such an array may also comprise replicates (e.g., duplicates, triplicates, etc.) of any one or more populations of host cells.

Arrays

Provided herein is an array of host cell populations (i.e. “strain array”) which can be rapidly screened to identify certain strain(s) with improved yield and/or quality of heterologous protein. As used herein, the term "strain array" refers to a plurality of addressed or addressable locations (e.g., wells, such as deep well or microwells). The location of each of the microwells or groups of microwells in the array is typically known, so as to allow for identification of the optimal host cell for expression of the heterologous protein of interest.

The strain array comprises a plurality of phenotypically distinct host strains. The arrays may be low-density arrays or high-density arrays and may contain about 2 or more, about 4 or more, about 8 or more, about 12 or more, about 16 or more, about 20 or more, about 24 or more, about 32 or more, about 40 or more, about 48 or more, about 64 or more, about 72 or more, about 80 or more, about 96 or more, about 192 or more, about 384 or more host cell populations.

The host cell populations of the invention can be maintained and/or screened in a multi-well or deep well vessel. The vessel may contain any desired number of wells, however, a miniaturized cell culture microarray platform is useful for screening each population of host cells individually and simultaneously using minimal reagents and a relatively small number of cells. A typical multi-well, microtiter vessel useful in this assay is a multi-well plate including, without limitation, 10-well plates, 28-well plates, 96-well plates, 384-well plates, and plates having greater than 384 wells. Alternatively, an array of tubes, holders, cartridges, minitubes, microfuge tubes, cryovials, square well plates tubes, plates, slants, or culture flasks may also be used, depending on the volume desired.

The vessel may be made of any material suitable for culturing and/or screening a host cell of interest, e.g., Pseudomonas. For example, the vessel can be a material that can be easily sterilized such as plastic or other artificial polymer material, so long as the material is biocompatible. Any number of materials can be used, including, but not limited to, polystyrene; polypropylene; polyvinyl compounds (e.g. polyvinylchloride); polycarbonate (PVC); polytetrafluoroethylene (PTFE); polyglycolic acid (PGA); cellulose; glass, fluoropolymers, fluorinated ethylene propylene, polyvinylidene, polydimethylsiloxane, silicon, and the like.

Automated transformation of cells and automated colony pickers will facilitate rapid screening of desired cells. The arrays may be created and/or screened using a spotter device (e.g., automated robotic devices) as known in the art.

Target genes

The strain array of the present invention comprises a plurality of phenotypically and genotypically distinct host cell populations, wherein each population in the array has been genetically modified to modulate the expression of one or more target genes in the host cell. By "target gene" is intended a gene that affects heterologous protein production in a host cell. A target gene may encode a host cell protease or an endogenous or exogenous protein folding modulator, transcription factor, translation factor, secretion modulator, or any other gene involved in the proper expression, processing, and/or translocation of a heterologous protein of interest. A "target protein" refers to the protein or polypeptide resulting from expression of the target gene. Expression and/or activity of a target gene or genes is increased or decreased, depending on the function of the target gene or protein. A target gene can be endogenous to the host cell, or can be a gene that is heterologously expressed in each of the host cell populations in the array.

In one embodiment, the target gene or genes is at least one protein folding modulator, putative protein folding modulator, or a cofactor or subunit of a folding modulator. In some embodiments, the target gene or genes can be selected from a chaperone protein, a foldase, a peptidyl prolyl isomerase and a disulfide bond isomerase. In some embodiments, the target gene or genes can be selected from htpG, cbpA, dnaJ, dnaK and fkbP. Exemplary protein folding modulators from *P. fluorescens* are listed in Table 1.

In other embodiments, the target gene comprises at least one putative protease, a protease-like protein, or a cofactor or subunit of a protease. For example, the target gene or genes can be a serine, threonine, cysteine, aspartic or metallopeptidase. In one embodiment, the target gene or genes can be selected from hslV, hslU, clpA, clpB and clpX. The target gene can also be a cofactor of a protease. Exemplary proteases from *P. fluorescens* are listed in Table 2. Proteases from a variety of organisms can be found in the MEROPS Peptidase Database maintained by the Wellcome Trust Sanger Institute, Cambridge, UK (see the website address merops.sanger.ac.uk/).

Protein folding modulators

Another major obstacle in the production of heterologous proteins in host cells is that the cell often is not adequately equipped to produce either soluble or active protein. While the primary structure of a protein is defined by its amino acid sequence, the secondary structure is defined by the presence of alpha helices or beta sheets, and the ternary structure by covalent bonds between adjacent protein stretches, such as disulfide bonds. When expressing heterologous proteins, particularly in large-scale production, the secondary and tertiary structure of the protein itself is of critical importance. Any significant change in protein structure can yield a functionally inactive molecule, or a protein with significantly reduced biological activity. In many cases, a host cell expresses protein folding modulators (PFMs) that are necessary for proper production of active heterologous protein. However, at the high levels of expression generally required to produce usable, economically satisfactory biotechnology products, a cell often cannot produce enough native protein folding modulator or modulators to process the heterologously-expressed protein.

In certain expression systems, overproduction of heterologous proteins can be accompanied by their misfolding and segregation into insoluble aggregates. In bacterial cells these aggregates are known as inclusion bodies. In *E. coli*, the network of folding modulators/chaperones includes the Hsp70 family. The major Hsp70 chaperone, DnaK, efficiently prevents protein aggregation and supports the refolding of damaged proteins. The incorporation of heat shock proteins into protein aggregates can facilitate disaggregation. However, proteins processed to inclusion bodies can, in certain cases, be recovered through additional processing of the insoluble fraction. Proteins found in inclusion bodies typically have to be purified through multiple steps, including denaturation and renaturation. Typical renaturation processes for

inclusion body targeted proteins involve attempts to dissolve the aggregate in concentrated denaturant and subsequent removal of the denaturant by dilution. Aggregates are frequently formed again in this stage. The additional processing adds cost, there is no guarantee that the *in vitro* refolding will yield biologically active product, and the recovered proteins can include large amounts of fragment impurities.

The recent realization that *in vivo* protein folding is assisted by molecular chaperones, which promote the proper isomerization and cellular targeting of other polypeptides by transiently interacting with folding intermediates, and by foldases, which accelerate rate-limiting steps along the folding pathway, has provided additional approaches to combat the problem of inclusion body formation (see for e.g. Thomas J G et al. (1997) *Appl Biochem Biotechnol* 66:197-238).

In certain cases, the overexpression of chaperones has been found to increase the soluble yields of aggregation-prone proteins (see Baneyx, F. (1999) *Curr. Opin. Biotech*. 10:411-421 and references therein). The beneficial effect associated with an increase in the intracellular concentration of these chaperones appears highly dependent on the nature of the overproduced protein, and may not require overexpression of the same protein folding modulator(s) for all heterologous proteins.

Protein folding modulators, including chaperones, disulfide bond isomerases, and peptidyl-prolyl cis-trans isomerases (PPIases) are a class of proteins present in all cells which aid in the folding, unfolding and degradation of nascent polypeptides.

Chaperones act by binding to nascent polypeptides, stabilizing them and allowing them to fold properly. Proteins possess both hydrophobic and hydrophilic residues, the former are usually exposed on the surface while the latter are buried within the structure where they interact with other hydrophilic residues rather than the water which surrounds the molecule. However in folding polypeptide chains, the hydrophilic residues are often exposed for some period of time as the protein exists in a partially folded or misfolded state. It is during this time when the forming polypeptides can become permanently misfolded or interact with other misfolded proteins and form large aggregates or inclusion bodies within the cell. Chaperones generally act by binding to the hydrophobic regions of the partially folded chains and preventing them from misfolding completely or aggregating with other proteins. Chaperones can even bind to proteins in inclusion bodies and allow them to

disaggregate (Ranson *et. al.* 1998). The GroES/EL, DnaKJ, Clp, Hsp90 and SecB families of folding modulators are all examples of proteins with chaperone like activity.

Another important type of folding modulator is the disulfide bond isomerases. These proteins catalyze a very specific set of reactions to help folding polypeptides form the proper intra-protein disulfide bonds. Any protein that has more than two cysteines is at risk of forming disulfide bonds between the wrong residues. The disulfide bond formation family consists of the Dsb proteins which catalyze the formation of disulfide bonds in the non-reducing environment of the periplasm. When a periplasmic polypeptide misfolds disulfide bond isomerase, DsbC is capable of rearranging the disulfide bonds and allowing the protein to reform with the correct linkages.

The proline residue is unique among amino acids in that the peptidyl bond immediately preceding it can adopt either a cis or trans conformation. For all other amino acids this is not favored due to steric hindrance. Peptidyl-prolyl cis-trans isomerases (PPIases) catalyze the conversion of this bond from one form to the other. This isomerization may aid in protein folding, refolding, assembly of subunits and trafficking in the cell (Dolinski, *et. al.*1997).

In addition to the general chaperones which seem to interact with proteins in a non-specific manner, there are also chaperones which aid in the folding of specific targets. These protein-specific chaperones form complexes with their targets, preventing aggregation and degradation and allowing time for them to assemble into multi-subunit structures. The PapD chaperone is one well known example of this type (Lombardo *et. al.* 1997).

Folding modulators also include, for example, HSP70 proteins, HSP110/SSE proteins, HSP40 (DNAJ-related) proteins, GRPE-like proteins, HSP90 proteins, CPN60 and CPN10 proteins, Cytosolic chaperoning, HSP100 proteins, Small HSPs, Calnexin and calreticulin, PDI and thioredoxin-related proteins, Peptidyl-prolyl isomerases, Cyclophilin PPIases, FK-506 binding proteins, Parvulin PPIases, Individual chaperoning, Protein specific chaperones, or intramolecular chaperones. Folding modulators are generally described in "Guidebook to Molecular Chaperones and Protein-Folding Catalysts" (1997) ed. M. Gething, Melbourne University, Australia.

The best characterized molecular chaperones in the cytoplasm of *E. coli* are the ATP-dependent DnaK-DnaJ-GrpE and GroEL-GroES systems. Based on *in vitro* studies and homology considerations, a number of additional cytoplasmic proteins have been proposed to function as molecular chaperones in *E. coli*. These include ClpB, HtpG and IbpA/B, which, like DnaK-DnaJ-GrpE and GroEL-GroES, are heat-shock proteins (Hsps) belonging to the stress regulon. The trans conformation of X-Pro bonds is energetically favored in nascent protein chains; however, approximately 5% of all prolyl peptide bonds are found in a cis conformation in native proteins. The trans to cis isomerization of X-Pro bonds is rate limiting in the folding of many polypeptides and is catalyzed *in vivo* by peptidyl prolyl cis/trans isomerases (PPIases). Three cytoplasmic PPIases, SlyD, SlpA and trigger factor (TF), have been identified to date in *E. coli*. TF, a 48 kDa protein associated with 50S ribosomal subunits that has been postulated to cooperate with chaperones in *E. coli* to guarantee proper folding of newly synthesized proteins. At least five proteins (thioredoxins 1 and 2, and glutaredoxins 1, 2 and 3, the products of the trxA, trxC, grxA, grxB and grxC genes, respectively) are involved in the reduction of disulfide bridges that transiently arise in cytoplasmic enzymes. Thus, target genes can be disulfide bond forming proteins or chaperones that allow proper disulfide bond formation.

Table 1. *P. fluorescens* strain MB214 protein folding modulators

ORF ID	GENE	FUNCTION	FAMILY	LOCATION
GroES/EL				
RXF02095.1	groES	Chaperone	Hsp10	Cytoplasmic
RXF06767.1::rxf02090	groEL	Chaperone	Hsp60	Cytoplasmic
RXF01748.1	ibpA	Small heat-shock protein (sHSP) IbpA PA3126;Acts as a holder for GroESL folding	Hsp20	Cytoplasmic
RXF03385.1	hscB	Chaperone protein hscB	Hsp20	Cytoplasmic
Hsp70 (DnaK/J)				
RXF05399.1	dnaK	Chaperone	Hsp70	Periplasmic
RXF06954.1	dnaK	Chaperone	Hsp70	Cytoplasmic
RXF03376.1	hscA	Chaperone	Hsp70	Cytoplasmic
RXF03987.2	cbpA	Curved dna-binding protein, dnaJ like activity	Hsp40	Cytoplasmic
RXF05406.2	dnaJ	Chaperone protein dnaJ	Hsp40	Cytoplasmic
RXF03346.2	dnaJ	Molecular chaperones (DnaJ family)	Hsp40	Non-secretory
RXF05413.1	grpE	heat shock protein GrpE PA4762	GrpE	Cytoplasmic
Hsp100 (Clp/Hsl)				
RXF04587.1	clpA	atp-dependent clp protease atp-binding subunit clpA	Hsp100	Cytoplasmic
RXF08347.1	clpB	ClpB protein	Hsp100	Cytoplasmic
RXF04654.2	clpX	atp-dependent clp protease atp-binding subunit clpX	Hsp100	Cytoplasmic
RXF04663.1	clpP	atp-dependent Clp protease proteolytic subunit (ec 3.4.21.92)	MEROPS peptidase family S14	Cytoplasmic
RXF01957.2	hslU	atp-dependent hsl protease atp-binding subunit hslU	Hsp100	Cytoplasmic
RXF01961.2	hslV	atp-dependent hsl protease proteolytic subunit	MEROPS peptidase subfamily T1B	Cytoplasmic
Hsp33				
RXF04254.2	yrfI	33 kDa chaperonin (Heat shock protein 33 homolog) (HSP33).	Hsp33	Cytoplasmic
Hsp90				
RXF05455.2	htpG	Chaperone protein htpG	Hsp90	Cytoplasmic
SecB				
RXF02231.1	secB	secretion specific chaperone SecB	SecB	Non-secretory
Disulfide Bond Isomerases				
RXF07017.2	dsbA	disulfide isomerase	DSBA oxido-reductase	Cytoplasmic

RXF08657.2	dsbA/ dsbC/ dsbG/ fernA	disulfide isomerase	DSBA oxido-reductase	Cytoplasmic
RXF01002.1	dsbA/ dsbC	disulfide isomerase	DSBA oxido-reductase/ Thioredoxin	Periplasmic
RXF03307.1	dsbC	disulfide isomerase	glutaredoxin/ Thioredoxin	Periplasmic
RXF04890.2	dsbG	disulfide isomerase	glutaredoxin/ Thioredoxin	Periplasmic
Peptidyl-prolyl cis-trans isomerases				
RXF03768.1	ppiA	Peptidyl-prolyl cis-trans isomerase A (ec 5.2.1.8)	PPIase: cyclophilin type	Periplasmic
RXF05345.2	ppiB	Peptidyl-prolyl cis-trans isomerase B.	PPIase: cyclophilin type	Cytoplasmic
RXF06034.2	fklB	Peptidyl-prolyl cis-trans isomerase FklB.	PPIase: FKBP type	OuterMembrane
RXF06591.1	fklB/ fkbP	fk506 binding protein Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8)	PPIase: FKBP type	Periplasmic
RXF05753.2	fklB; fkbP	Peptidyl-prolyl cis-trans isomerase (ec 5.2.1.8)	PPIase: FKBP type	Outer Membrane
RXF01833.2	slyD	Peptidyl-prolyl cis-trans isomerase SlyD.	PPIase: FKBP type	Non-secretory
RXF04655.2	tig	Trigger factor, ppiase (ec 5.2.1.8)	PPIase: FKBP type	Cytoplasmic
RXF05385.1	yaad	Probable FKBP-type 16 kDa peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPiase) (Rotamase).	PPIase: FKBP type	Non-secretory
RXF00271.1		Peptidyl-prolyl cis-trans isomerase (ec 5.2.1.8)	PPIase: FKBP type	Non-secretory
pili assembly chaperones (papD like)				
RXF06068.1	cup	Chaperone protein cup	pili assembly papD	Periplasmic
RXF05719.1	ecpD	Chaperone protein ecpD	pili assembly papD	Signal peptide
RXF03406.2	ecpD; csuC	Chaperone protein ecpD	pili assembly papD	Signal peptide
RXF04296.1	ecpD; cup	Chaperone protein ecpD	pili assembly papD	Periplasmic
RXF04553.1	ecpD; cup	Chaperone protein ecpD	pili assembly papD	Periplasmic
RXF04554.2	ecpD; cup	Chaperone protein ecpD	pili assembly papD	Periplasmic
RXF05310.2	ecpD; cup	Chaperone protein ecpD	pili assembly papD	Periplasmic
RXF05304.1	ecpD; cup	Chaperone protein ecpD	pili assembly papD	Periplasmic
RXF05073.1	gltF	Gram-negative pili assembly chaperone periplasmic function	pili assembly papD	Signal peptide

Protease

Unwanted degradation of heterologously-expressed protein presents an obstacle to the efficient use of certain expression systems. When a cell is modified to produce large quantities of a target protein, the cell is placed under stress and often reacts by inducing or suppressing other proteins. The stress that a host cell undergoes during production of heterologous proteins can increase expression of, for example, specific proteins or cofactors to cause degradation of the overexpressed heterologous protein. The increased expression of compensatory proteins can be counterproductive to the goal of expressing high levels of active, full-length heterologous protein. Decreased expression or lack of adequate expression of other proteins can cause misfolding and aggregation of the heterologously-expressed protein. While it is known that a cell under stress will change its profile of protein expression, not all heterologously expressed proteins will modulate expression of the same proteins in a particular host cell.

Thus, the optimal host strain, e.g., *P. fluorescens* host strain, can be identified using an array comprising a plurality of host cell populations that have been genetically engineered to decrease the expression of one or more protease enzymes. In one embodiment, one or more host cell populations is modified by reducing the expression of, inhibiting or removing at least one protease from the genome. The modification can also be to more than one protease. In a related embodiment, the cell is modified by reducing the expression of a protease cofactor or protease protein. In another embodiment, the host cell is modified by inhibition of a promoter for a protease or related protein, which can be a native promoter. Alternatively, the gene modification can be to modulate a protein homologous to the target gene.

The array comprising the modified host strains can be screened by expressing the heterologous protein(s) of interest and assessing the quality and/or quantity of protein production as discussed *infra*. Alternatively, an isolate of the heterologous protein of interest can be independently incubated with lysate collected from each of the protease-deficient host cell populations and the level of proteolytic degradation can be used to identify the optimal host cell. In this embodiment, the optimal host cell population is that which results in the least amount of heterologous protein degradation. Thus, in one embodiment, lysate from the optimal host cell population can be degraded by less than about 50% of the heterologous protein, less than about

45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 10%, less than about 5%, less than about 4%, about 3%, about 2%, about 1%, or less of the protein.

Exemplary target protease genes include those proteases classified as Aminopeptidases; Dipeptidases; Dipeptidyl-peptidases and tripeptidyl peptidases; Peptidyl-dipeptidases; Serine-type carboxypeptidases; Metallocarboxypeptidases; Cysteine-type carboxypeptidases; Omegapeptidases; Serine proteinases; Cysteine proteinases; Aspartic proteinases; Metallo proteinases; or Proteinases of unknown mechanism.

Aminopeptidases include cytosol aminopeptidase (leucyl aminopeptidase), membrane alanyl aminopeptidase, cystinyl aminopeptidase, tripeptide aminopeptidase, prolyl aminopeptidase, arginyl aminopeptidase, glutamyl aminopeptidase, x-pro aminopeptidase, bacterial leucyl aminopeptidase, thermophilic aminopeptidase, clostridial aminopeptidase, cytosol alanyl aminopeptidase, lysyl aminopeptidase, x-trp aminopeptidase, tryptophanyl aminopeptidase, methionyl aminopeptidas, d-stereospecific aminopeptidase, aminopeptidase ey. Dipeptidases include x-his dipeptidase, x-arg dipeptidase, x-methyl-his dipeptidase, cys-gly dipeptidase, glu-glu dipeptidase, pro-x dipeptidase, x-pro dipeptidase, met-x dipeptidase, non-stereospecific dipeptidase, cytosol non-specific dipeptidase, membrane dipeptidase, beta-ala-his dipeptidase. Dipeptidyl-peptidases and tripeptidyl peptidases include dipeptidyl-peptidase i, dipeptidyl-peptidase ii, dipeptidyl peptidase iii, dipeptidyl-peptidase iv, dipeptidyl-dipeptidase, tripeptidyl-peptidase I, tripeptidyl-peptidase II. Peptidyl-dipeptidases include peptidyl-dipeptidase a and peptidyl-dipeptidase b. Serine-type carboxypeptidases include lysosomal pro-x carboxypeptidase, serine-type D-ala-D-ala carboxypeptidase, carboxypeptidase C, carboxypeptidase D. Metallocarboxypeptidases include carboxypeptidase a, carboxypeptidase B, lysine(arginine) carboxypeptidase, gly-X carboxypeptidase, alanine carboxypeptidase, muramoylpentapeptide carboxypeptidase, carboxypeptidase h, glutamate carboxypeptidase, carboxypeptidase M, muramoyltetrapeptide carboxypeptidase, zinc d-ala-d-ala carboxypeptidase, carboxypeptidase A2, membrane pro-x carboxypeptidase, tubulinyl-tyr carboxypeptidase, carboxypeptidase t. Omegapeptidases include acylaminoacyl-peptidase, peptidyl-glycinamidase, pyroglutamyl-peptidase I, beta-aspartyl-peptidase, pyroglutamyl-peptidase II, n-formylmethionyl-peptidase, pteroylpoly-[gamma]-glutamate carboxypeptidase,

gamma-glu-X carboxypeptidase, acylmuramoyl-ala peptidase. Serine proteinases include chymotrypsin, chymotrypsin c, metridin, trypsin, thrombin, coagulation factor Xa, plasmin, enteropeptidase, acrosin, alpha-lytic protease, glutamyl, endopeptidase, cathepsin G, coagulation factor viia, coagulation factor ixa, cucumisi, prolyl oligopeptidase, coagulation factor xia, brachyurin, plasma kallikrein, tissue kallikrein, pancreatic elastase, leukocyte elastase, coagulation factor xiia, chymase, complement component c1r55, complement component c1s55, classical-complement pathway c3/c5 convertase, complement factor I, complement factor D, alternative-complement pathway c3/c5 convertase, cerevisin, hypodermin C, lysyl endopeptidase, endopeptidase 1a, gamma-reni, venombin ab, leucyl endopeptidase, tryptase, scutelarin, kexin, subtilisin, oryzin, endopeptidase k, thermomycolin, thermitase, endopeptidase SO, T-plasminogen activator, protein C, pancreatic endopeptidase E, pancreatic elastase ii, IGA-specific serine endopeptidase, U-plasminogen, activator, venombin A, furin, myeloblastin, semenogelase, granzyme A or cytotoxic T-lymphocyte proteinase 1, granzyme B or cytotoxic T-lymphocyte proteinase 2, streptogrisin A, treptogrisin B, glutamyl endopeptidase II, oligopeptidase B, limulus clotting factor c, limulus clotting factor, limulus clotting enzyme, omptin, repressor lexa, bacterial leader peptidase I, togavirin, flavirin. Cysteine proteinases include cathepsin B, papain, ficin, chymopapain, asclepain, clostripain, streptopain, actinide, cathepsin 1, cathepsin H, calpain, cathepsin t, glycyl, endopeptidase, cancer procoagulant, cathepsin S, picornain 3C, picornain 2A, caricain, ananain, stem bromelain, fruit bromelain, legumain, histolysain, interleukin 1-beta converting enzyme. Aspartic proteinases include pepsin A, pepsin B, gastricsin, chymosin, cathepsin D, neopenthesin, renin, retropepsin, pro-opiomelanocortin converting enzyme, aspergillopepsin I, aspergillopepsin II, penicillopepsin, rhizopuspepsin, endothiapepsin, mucoropepsin, candidapepsin, saccharopepsin, rhodotorulapepsin, physaropepsin, acrocylindropepsin, polyporopepsin, pycnoporopepsin, scytalidopepsin a, scytalidopepsin b, xanthomonapepsin, cathepsin e, barrierpepsin, bacterial leader peptidase I, pseudomonapepsin, plasmepsin. Metallo proteinases include atrolysin a, microbial collagenase, leucolysin, interstitial collagenase, neprilysin, envelysin, iga-specific metalloendopeptidase, procollagen N-endopeptidase, thimet oligopeptidase, neurolysin, stromelysin 1, meprin A, procollagen C-endopeptidase, peptidyl-lys metalloendopeptidase, astacin, stromelysin, 2, matrilysin gelatinase, aeromonolysin, pseudolysin, thermolysin,

bacillolysin, aureolysin, coccolysin, mycolysin, beta-lytic metalloendopeptidase, peptidyl-asp metalloendopeptidase, neutrophil collagenase, gelatinase B, leishmanolysin, saccharolysin, autolysin, deuterolysin, serralysin, atrolysin B, atrolysin C, atroxase, atrolysin E, atrolysin F, adamalysin, horrilysin, ruberlysin, bothropasin, bothrolysin, ophiolysin, trimerelysin I, trimerelysin II, mucrolysin, pitrilysin, insulysin, O-syaloglycoprotein endopeptidase, russellysin, mitochondrial, intermediate, peptidase, dactylysin, nardilysin, magnolysin, meprin B, mitochondrial processing peptidase, macrophage elastase, choriolysin, toxilysin. Proteinases of unknown mechanism include thermopsin and multicatalytic endopeptidase complex.

Table 2. *P. fluorescens* strain MB214 proteases

Class	Family	RXF	Gene	Curated Function	Location
Aspartic Peptidases					
	A8 (signal peptidase II family)				
		RXF05383.2		Lipoprotein signal peptidase (ec 3.4.23.36)	Cytoplasmic Membrane
	A24 (type IV prepilin peptidase family)				
		RXF05379.1		type 4 prepilin peptidase pild (ec 3.4.99.-)	Cytoplasmic Membrane
Cysteine Peptidases					
	C15 (pyroglutamyl peptidase I family)				
		RXF02161.1		Pyrrolidone-carboxylate peptidase (ec 3.4.19.3)	Cytoplasmic
	C40				
		RXF01968.1		invasion-associated protein, P60	Signal peptide
		RXF04920.1		invasion-associated protein, P60	Cytoplasmic
		RXF04923.1		phosphatase-associated protein papq	Signal peptide
	C56 (PfpI endopeptidase family)				
		RXF01816.1		protease I (ec 3.4.-.-)	Non-secretory
Metallopeptidases					
	M1				
		RXF08773.1		Membrane alanine aminopeptidase (ec 3.4.11.2)	Non-secretory
	M3				
		RXF00561.2	prlC	Oligopeptidase A (ec 3.4.24.70)	Cytoplasmic
		RXF04631.2		Zn-dependent oligopeptidases	Cytoplasmic
	M4 (thermolysin family)				
		RXF05113.2		Extracellular metalloprotease precursor (ec 3.4.24.-)	Extracellular
	M41 (FtsH endopeptidase family)				
		RXF05400.2		Cell division protein ftsH (ec 3.4.24.-)	Cytoplasmic Membrane
	M10				
		RXF04304.1		Serralysin (ec 3.4.24.40)	Extracellular
		RXF04500.1		Serralysin (ec 3.4.24.40)	Extracellular
		RXF01590.2		Serralysin (ec 3.4.24.40)	Extracellular
		RXF04495.2		Serralysin (ec 3.4.24.40)	Extracellular
		RXF02796.1		Serralysin (ec 3.4.24.40)	Extracellular
	M14 (carboxypeptidase A family)				
		RXF09091.1		Zinc-carboxypeptidase precursor (ec 3.4.17.-)	Cytoplasmic
	M16 (pitrilysin family)				
		RXF03441.1		Coenzyme pqq synthesis protein F (ec 3.4.99.-)	Non-secretory

	RXF01918.1	zinc protease (ec 3.4.99.-)	Signal peptide
	RXF01919.1	zinc protease (ec 3.4.99.-)	Periplasmic
	RXF03699.2	processing peptidase (ec 3.4.24.64)	Signal peptide
M17 (leucyl aminopeptidase family)			
	RXF00285.2	Cytosol aminopeptidase (ec 3.4.11.1)	Non-secretory
M18			
	RXF07879.1	Aspartyl aminopeptidase (ec 3.4.11.21)	Cytoplasmic
M20			
	RXF00811.1 dapE	Succinyl-diaminopimelate desuccinylase (ec 3.5.1.18)	Cytoplasmic
	RXF04052.2	Xaa-His dipeptidase (ec 3.4.13.3)	Signal peptide
	RXF01822.2	Carboxypeptidase G2 precursor (ec 3.4.17.11)	Signal peptide
	RXF09831.2:: RXF04892.1	N-acyl-L-amino acid amidohydrolase (ec 3.5.1.14)	Signal peptide
M28 (aminopeptidase Y family)			
	RXF03488.2	Alkaline phosphatase isozyme conversion protein precursor (ec 3.4.11.-)	OuterMembrane
M42 (glutamyl aminopeptidase family)			
	RXF05615.1	Deblocking aminopeptidase (ec 3.4.11.-)	Non-secretory
M22			
	RXF05817.1	O-sialoglycoprotein endopeptidase (ec 3.4.24.57)	Extracellular
	RXF03065.2	Glycoprotease protein family	Non-secretory
M23			
	RXF01291.2	Cell wall endopeptidase, family M23/M37	Signal peptide
	RXF03916.1	Membrane proteins related to metalloendopeptidases	Signal peptide
	RXF09147.2	Cell wall endopeptidase, family M23/M37	Signal peptide
M24			
	RXF04693.1	Methionine aminopeptidase (ec 3.4.11.18)	Cytoplasmic
	RXF03364.1	Methionine aminopeptidase (ec 3.4.11.18)	Non-secretory
	RXF02980.1	Xaa-Pro aminopeptidase (ec 3.4.11.9)	Cytoplasmic
	RXF06564.1	Xaa-Pro aminopeptidase (ec 3.4.11.9)	Cytoplasmic
M48 (Ste24 endopeptidase family)			
	RXF05137.1	Heat shock protein HtpX	Cytoplasmic Membrane
	RXF05081.1	Zinc metalloprotease (ec 3.4.24.-)	Signal peptide

M50 (S2P protease family)			
	RXF04692.1	Membrane metalloprotease	Cytoplasmic Membrane
Serine Peptidases			
S1 (chymotrypsin family)			
	RXF01250.2	protease do (ec 3.4.21.-)	Periplasmic
	RXF07210.1	protease do (ec 3.4.21.-)	Periplasmic
S8 (subtilisin family)			
	RXF06755.2	serine protease (ec 3.4.21.-)	Non-secretory
	RXF08517.1	serine protease (ec 3.4.21.-)	Extracellular
	RXF08627.2	extracellular serine protease (ec 3.4.21.-)	Signal peptide
	RXF06281.1	Extracellular serine protease precursor (ec 3.4.21.-)	Non-secretory
	RXF08978.1	extracellular serine protease (ec 3.4.21.-)	OuterMembrane
	RXF06451.1	serine protease (ec 3.4.21.-)	Signal peptide
S9 (prolyl oligopeptidase family)			
	RXF02003.2	Protease ii (ec 3.4.21.83)	Periplasmic
	RXF00458.2	Hydrolase	Non-secretory
S11 (D-Ala-D-Ala carboxypeptidase A family)			
	RXF04657.2	D-alanyl-D-alanine-endopeptidase (ec 3.4.99.-)	Periplasmic
	RXF00670.1	D-alanyl-D-alanine carboxypeptidase (ec 3.4.16.4)	Cytoplasmic Membrane
S13 (D-Ala-D-Ala peptidase C family)			
	RXF00133.1	D-alanyl-meso-diaminopimelate endopeptidase (ec 3.4.-.-)	OuterMembrane
	RXF04960.2	D-alanyl-meso-diaminopimelate endopeptidase (ec 3.4.-.-)	Signal peptide
S14 (ClpP endopeptidase family)			
	RXF04567.1 clpP	atp-dependent Clp protease proteolytic subunit (ec 3.4.21.92)	Non-secretory
	RXF04663.1 clpP	atp-dependent Clp protease proteolytic subunit (ec 3.4.21.92)	Cytoplasmic
S16 (lon protease family)			
	RXF04653.2	atp-dependent protease La (ec 3.4.21.53)	Cytoplasmic
	RXF08653.1	atp-dependent protease La (ec 3.4.21.53)	Cytoplasmic
	RXF05943.1	atp-dependent protease La (ec 3.4.21.53)	Cytoplasmic
S24 (LexA family)			
	RXF00449.1	LexA repressor (ec 3.4.21.88)	Non-secretory
	RXF03397.1	LexA repressor (ec 3.4.21.88)	Cytoplasmic

S26 (signal peptidase I family)			
	RXF01181.1	Signal peptidase I (ec 3.4.21.89)	Cytoplasmic Membrane
S33			
	RXF05236.1 pip3	Proline iminopeptidase (ec 3.4.11.5)	Non-secretory
	RXF04802.1 pip1	Proline iminopeptidase (ec 3.4.11.5)	Non-secretory
	RXF04808.2 pip2	Proline iminopeptidase (ec 3.4.11.5)	Cytoplasmic
S41 (C-terminal processing peptidase family)			
	RXF06586.1	Tail-specific protease (ec 3.4.21.-)	Signal peptide
	RXF01037.1	Tail-specific protease (ec 3.4.21.-)	Signal peptide
S45			
	RXF07170.1 pacB 2	Penicillin acylase (ec 3.5.1.11)	Signal peptide
	RXF06399.2 pacB 1	Penicillin acylase ii (ec 3.5.1.11)	Signal peptide
S49 (protease IV family)			
	RXF06993.2	possible protease sohb (ec 3.4.-.-)	Non-secretory
	RXF01418.1	protease iv (ec 3.4.-.-)	Non-secretory
S58 (DmpA aminopeptidase family)			
	RXF06308.2	D-aminopeptidase (ec 3.4.11.19)	Cytoplasmic Membrane
Threonine Peptidases			
T1 (proteasome family)			
	RXF01961.2 hslV	atp-dependent protease hslV (ec 3.4.25.-)	Cytoplasmic
T3 (gamma-glutamyltransferase family)			
	RXF02342.1 ggt1	Gamma-glutamyltranspeptidase (ec 2.3.2.2)	Periplasmic
	RXF04424.2 ggt2	Gamma-glutamyltranspeptidase (ec 2.3.2.2)	Periplasmic
Unclassified Peptidases			
U32			
	RXF00428.1	protease (ec 3.4.-.-)	Cytoplasmic
	RXF02151.2	protease (ec 3.4.-.-)	Cytoplasmic
U61			
	RXF04715.1	Muramoyltetrapeptide carboxypeptidase (ec 3.4.17.13)	Non-secretory
U62			
	RXF04971.2 pmbA	PmbA protein	Cytoplasmic
	RXF04968.2	TldD protein	Cytoplasmic
Non MEROPS Proteases			
	RXF00325.1	Repressor protein C2	Non-secretory
	RXF02689.2	Microsomal dipeptidase (ec 3.4.13.19)	Cytoplasmic
	RXF02739.1	membrane dipeptidase (3.4.13.19)	Signal peptide
	RXF03329.2	Hypothetical Cytosolic Protein	Cytoplasmic

	RXF02492.1	Xaa-Pro dipeptidase (ec 3.4.13.9)	Cytoplasmic
	RXF04047.2	caax amino terminal protease family	Cytoplasmic Membrane
	RXF08136.2	protease (transglutaminase-like protein)	Cytoplasmic
	RXF09487.1	Zinc metalloprotease (ec 3.4.24.-)	Non-secretory

Additional protein modification enzymes

In another embodiment, the target gene comprises a gene involved in proper protein processing and/or modification. Common modifications include disulfide bond formation, glycosylation, acetylation, acylation, phosphorylation, and gamma-carboxylation, all of which can regulate protein folding and biological activity. A non-exhaustive list of several classes of enzymes involved in protein processing is found in Table 3. One of skill in the art will recognize how to identify a target gene useful in the host cell chosen for the array, or useful with the heterologous protein of interest, from among the classes of protein modification enzymes listed in Table 3. The target gene may be endogenous to the host cell utilized, may be endogenous to the organism from which the heterologous protein of interest is derived, or may be known to facilitate proper processing of a heterologously expressed protein of interest. It is also recognized that any gene involved in protein production can be targeted according to desired specifications for the heterologous protein of interest.

Table 3. Classes of enzymes involved in protein processing

Class	Examples
Glycosyltransferases (EC 2.4.1.18)	
	α-glucan-branching glycosyltransferase
	enzymatic branching factor
	branching glycosyltransferase
	enzyme Q
	glucosan transglycosylase
	glycogen branching enzyme
	amylose isomerase
	plant branching enzyme
	α-1,4-glucan:α-1,4-glucan-6-glycosyltransferase
	starch branching enzyme
	UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase
	GDP-fucose protein O-fucosyltransferase 2
	O-GlcNAc transferase

Histone acetyltransferase (EC 2.3.1.48)
nucleosome-histone acetyltransferase
histone acetokinase
histone acetylase
histone transacetylase
histone deacetylase
Protein kinase (EC 2.7)
non-specific serine/threonine protein kinase
Fas-activated serine/threonine kinase
Goodpasture antigen-binding protein kinase
IκB kinase
cAMP-dependent protein kinase
cGMP-dependent protein kinase
protein kinase C
polo kinase
cyclin-dependent kinase
mitogen-activated protein kinase
mitogen-activated protein kinase kinase kinase
receptor protein serine/threonine kinase
dual-specificity kinase
Phosphatase (EC 3.1.3.48)
protein-tyrosine-phosphatase
phosphotyrosine phosphatase
phosphoprotein phosphatase (phosphotyrosine)
phosphotyrosine histone phosphatase
protein phosphotyrosine phosphatase
tyrosylprotein phosphatase
phosphotyrosine protein phosphatase
phosphotyrosylprotein phosphatase
tyrosine O-phosphate phosphatase
PPT-phosphatase
PTPase
[phosphotyrosine]protein phosphatase
PTP-phosphatase

Methods for modulating the expression of target genes

One or more host cell populations of the array can be modified by any technique known in the art, for example by a technique wherein at least one target gene is knocked out of the genome, or by mutating at least one target gene to reduce expression of the gene, by altering at least one promoter of at least one target gene to reduce expression of the target gene, or by coexpressing (with the heterologous protein or polypeptide of interest) the target gene or an inhibitor of the target gene in

the host genome. As discussed *supra*, the target gene can be endogenous to the host cell populations in the array, or can be heterologously expressed in each of the host cell populations.

The expression of target genes can be increased, for example, by introducing into at least one cell in a host population an expression vector comprising one or more target genes involved in protein production. The target gene expression can also be increased, for example, by mutating a promoter of a target gene. A host cell or organism that expresses a heterologous protein can also be genetically modified to increase the expression of at least one target gene involved in protein production and decrease the expression of at least one target gene involved in protein degradation.

The genome may be modified to modulate the expression of one or more target genes by including an exogenous gene or promoter element in the genome or in the host with an expression vector, by enhancing the capacity of a particular target gene to produce mRNA or protein, by deleting or disrupting a target gene or promoter element, or by reducing the capacity of a target gene to produce mRNA or protein. The genetic code can be altered, thereby affecting transcription and/or translation of a target gene, for example through substitution, deletion ("knock-out"), co-expression, or insertion ("knock-in") techniques. Additional genes for a desired protein or regulatory sequence that modulate transcription of an existing target sequence can also be inserted.

Genome modification

The genome of the host cell can be modified via a genetic targeting event, which can be by insertion or recombination, for example homologous recombination. Homologous recombination refers to the process of DNA recombination based on sequence homology. Homologous recombination permits site-specific modifications in endogenous genes and thus novel alterations can be engineered into a genome (see, for example Radding (1982) *Ann. Rev. Genet*. 16: 405; U.S. Pat. No. 4,888,274).

Various constructs can be prepared for homologous recombination at a target locus. Usually, the construct can include at least 10 bp, 20 bp, 30 bp, 40 bp, 50 bp, 70 bp, 100 bp, 500 bp, 1 kbp, 2 kbp, 4 kbp, 5 kbp, 10 kbp, 15 kbp, 20 kbp, or 50 kbp of sequence homologous with the identified locus. Various considerations can be involved in determining the extent of homology of target gene sequences, such as, for

example, the size of the target locus, availability of sequences, relative efficiency of double cross-over events at the target locus and the similarity of the target sequence with other sequences.

The modified gene can include a sequence in which DNA substantially isogenic flanks the desired sequence modifications with a corresponding target sequence in the genome to be modified. The "modified gene" is the sequence being introduced into the genome to alter the expression of a protease or a protein folding modulator in the host cell. The "target gene" is the sequence that is being replaced by the modified gene. The substantially isogenic sequence can be at least about 95%, 97-98%, 99.0-99.5%, 99.6-99.9%, or 100% identical to the corresponding target sequence (except for the desired sequence modifications). The modified gene and the targeted gene can share stretches of DNA at least about 10, 20, 30, 50, 75, 150 or 500 base pairs that are 100% identical.

Nucleotide constructs can be designed to modify the endogenous, target gene product. The modified gene sequence can have one or more deletions, insertions, substitutions or combinations thereof designed to disrupt the function of the resultant gene product. In one embodiment, the alteration can be the insertion of a selectable marker gene fused in reading frame with the upstream sequence of the target gene.

The genome can also be modified using insertional inactivation. In this embodiment, the genome is modified by recombining a sequence in the gene that inhibits gene product formation. This insertion can either disrupt the gene by inserting a separate element, or remove an essential portion of the gene. In one embodiment, the insertional deletion also includes insertion of a gene coding for resistance to a particular stressor, such as an antibiotic, or for growth in a particular media, for example for production of an essential amino acid.

The genome can also be modified by use of transposons, which are genetic elements capable of inserting at sites in prokaryote genomes by mechanisms independant of homologous recombination. Transposons can include, for example, Tn7, Tn5, or Tn10 in E. coli, Tn554 in S. aureus, IS900 in M. paratuberculosis, IS492 from Pseudomonas atlantica, IS116 from Streptomyces and IS900 from M. paratuberculosis. Steps believed to be involved in transposition include cleavage of the end of the transposon to yield 3' OH; strand transfer, in which transposase brings together the 3'OH exposed end of transposon and the identified sequence; and a single

step transesterification reaction to yield a covalent linkage of the transposon to the identified DNA. The key reaction performed by transposase is generally thought to be nicking or strand exchange, the rest of the process is done by host enzymes.

In one embodiment, the expression or activity of a target gene or protein is increased by incorporating a genetic sequence encoding the target protein or homolog thereof into the genome by recombination. In another embodiment, a promoter is inserted into the genome to enhance the expression of the target gene or homolog. In another embodiment, the expression or activity of a target gene or homolog thereof is decreased by recombination with an inactive gene. In another embodiment, a sequence that encodes a different gene, which can have a separate function in the cell or can be a reporter gene such as a resistance marker or an otherwise detectable marker gene, can be inserted into the genome through recombination. In yet another embodiment, a copy of at least a portion of the target gene that has been mutated at one or more locations is inserted into the genome through recombination. The mutated version of the target gene may not encode a protein, or the protein encoded by the mutated gene may be rendered inactive, the activity may be modulated (either increased or decreased), or the mutant protein can have a different activity when compared to the native protein.

There are strategies to knock out genes in bacteria, which have been generally exemplified in *E. coli*. One route is to clone a gene-internal DNA fragment into a vector containing an antibiotic resistance gene (e.g. ampicillin). Before cells are transformed via conjugative transfer, chemical transformation or electroporation (Puehler, et al. (1984) Advanced Molecular Genetics New York, Heidelberg, Berlin, Tokyo, Springer Verlag), an origin of replication, such as the vegetative plasmid replication (the oriV locus) is excised and the remaining DNA fragment is re-ligated and purified (Sambrook, et al. (2000) Molecular cloning: A laboratory manual, third edition Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press). Alternatively, antibiotic-resistant plasmids that have a DNA replication origin can be used. After transformation, the cells are plated onto e.g. LB agar plates containing the appropriate antibiotics (e.g. 200 micrograms/mL ampicillin). Colonies that grow on the plates containing the antibiotics presumably have undergone a single recombination event (Snyder, L., W. Champness, et al. (1997) Molecular Genetics of Bacteria Washington DC, ASM Press) that leads to the integration of the entire DNA fragment into the genome at the homologous locus. Further analysis of the antibiotic-

resistant cells to verify that the desired gene knock-out has occurred at the desired locus is e.g. by diagnostic PCR (McPherson, M. J., P. Quirke, et al. (1991) PCR: A Practical Approach New York, Oxford University Press). Here, at least two PCR primers are designed: one that hybridizes outside the DNA region that was used for the construction of the gene knock-out; and one that hybridizes within the remaining plasmid backbone. Successful PCR amplification of the DNA fragment with the correct size followed by DNA sequence analysis will verify that the gene knock-out has occurred at the correct location in the bacterial chromosome. The phenotype of the newly constructed mutant strain can then be analyzed by, e.g., SDS polyacrylamide gel electrophoresis (Simpson, R. J. (2003) Proteins and Proteomics--A Laboratory Manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).

An alternate route to generate a gene knock-out is by use of a temperature-sensitive replicon, such as the pSC101 replicon to facilitate gene replacement (Hamilton, et al. (1989) *Journal of Bacteriology* 171(9): 4617-22). The process proceeds by homologous recombination between a gene on a chromosome and homologous sequences carried on a plasmid temperature sensitive for DNA replication. After transformation of the plasmid into the appropriate host, it is possible to select for integration of the plasmid into the chromosome at 44° C. Subsequent growth of these cointegrates at 30° C leads to a second recombination event, resulting in their resolution. Depending on where the second recombination event takes place, the chromosome will either have undergone a gene replacement or retain the original copy of the gene.

Other strategies have been developed to inhibit expression of particular gene products. For example, RNA interference (RNAi), particularly using small interfering RNA (siRNA), has been extensively developed to reduce or even eliminate expression of a particular gene product. siRNAs are short, double-stranded RNA molecules that can target complementary mRNAs for degradation. RNAi is the phenomenon in which introduction of a double-stranded RNA suppresses the expression of the homologous gene. dsRNA molecules are reduced *in vivo* to 21-23 nt siRNAs which are the mediators of the RNAi effect. Upon introduction, double stranded RNAs get processed into 20-25 nucleotide siRNAs by an RNase III-like enzyme called Dicer (initiation step). Then, the siRNAs assemble into endoribonuclease-containing complexes known as RNA-induced silencing complexes (RISCs), unwinding in the

process. The siRNA strands subsequently guide the RISCs to complementary RNA molecules, where they cleave and destroy the cognate RNA (effecter step). Cleavage of cognate RNA takes place near the middle of the region bound by the siRNA strand. RNAi has been successfully used to reduce gene expression in a variety of organisms including zebrafish, nematodes (C. elegans), insects (Drosophila melanogaster), planaria, cnidaria, trypanosomes, mice and mammalian cells.

The genome can also be modified by mutation of one or more nucleotides in an open reading frame encoding a target gene. Techniques for genetic mutation, for instance site directed mutagenesis, are well known in the art. Some approaches focus on the generation of random mutations in chromosomal DNA such as those induced by X-rays and chemicals.

Coexpression

In one embodiment, one or more target genes in the host cell can be modified by including one or more vectors that encode the target gene(s) to facilitate coexpression of the target gene with the heterologous protein or peptide. In another embodiment, the host cell is modified by enhancing a promoter for a target gene, including by adding an exogenous promoter to the host cell genome.

In another embodiment, one or more target genes in the host cell is modified by including one or more vectors that encode an inhibitor of a target gene, such as a protease inhibitor to inhibit the activity of a target protease. Such an inhibitor can be an antisense molecule that limits the expression of the target gene, a cofactor of the target gene or a homolog of the target gene. Antisense is generally used to refer to a nucleic acid molecule with a sequence complementary to at least a portion of the target gene. In addition, the inhibitor can be an interfering RNA or a gene that encodes an interfering RNA. In Eukaryotic organisms, such an interfering RNA can be a small interfering RNA or a ribozyme, as described, for example, in Fire, A. et al. (1998) Nature 391:806-11, Elbashir et al. (2001) Genes & Development 15(2):188-200, Elbashir et al. (2001) Nature 411(6836):494-8, U.S. Pat. Nos. 6,506,559 to Carnegie Institute, 6,573,099 to Benitec, U.S. patent application Nos. 2003/0108923 to the Whitehead Inst., and 2003/0114409, PCT Publication Nos. WO03/006477, WO03/012052, WO03/023015, WO03/056022, WO03/064621 and WO03/070966.

The inhibitor can also be another protein or peptide. The inhibitor can, for example, be a peptide with a consensus sequence for the target protein. The inhibitor can also be a protein or peptide that can produce a direct or indirect inhibitory molecule for the target protein in the host. For example, protease inhibitors can include Amastatin, E-64, Antipain, Elastatinal, APMSF, Leupeptin, Bestatin, Pepstatin, Benzamidine, 1,10-Phenanthroline, Chymostatin, Phosphoramidon, 3,4-dichloroisocoumarin, TLCK, DFP, TPCK. Over 100 naturally occurring protein protease inhibitors have been identified so far. They have been isolated in a variety of organisms from bacteria to animals and plants. They behave as tight-binding reversible or pseudo-irreversible inhibitors of proteases preventing substrate access to the active site through steric hindrance. Their size are also extremely variable from 50 residues (e.g BPTI: Bovine Pancreatic Trypsin Inhibitor) to up to 400 residues (e.g alpha-1PI: alpha-1 Proteinase Inhibitor). They are strictly class-specific except proteins of the alpha-macroglobulin family (e.g alpha-2 macroglobulin) which bind and inhibit most proteases through a molecular trap mechanism.

An exogenous vector or DNA construct can be transfected or transformed into the host cell. Techniques for transfecting and transforming eukaryotic and prokaryotic cells respectively with exogenous nucleic acids are well known in the art. These can include lipid vesicle mediated uptake, calcium phosphate mediated transfection (calcium phosphate/DNA co-precipitation), viral infection, particularly using modified viruses such as, for example, modified adenoviruses, microinjection and electroporation. For prokaryotic transformation, techniques can include heat shock mediated uptake, bacterial protoplast fusion with intact cells, microinjection and electroporation. Techniques for plant transformation include Agrobacterium mediated transfer, such as by *A. tumefaciens*, rapidly propelled tungsten or gold microprojectiles, electroporation, microinjection and polyethelyne glycol mediated uptake. The DNA can be single or double stranded, linear or circular, relaxed or supercoiled DNA. For various techniques for transfecting mammalian cells, see, for example, Keown et al. (1990) Processes in Enzymology Vol. 185, pp. 527-537.

An expression construct encoding a target gene or an enhancer or inhibitor thereof can be constructed as described below for the expression constructs comprising the heterologous protein or polypeptide of interest. For example, the constructs can contain one, or more than one, internal ribosome entry site (IRES). The construct can also contain a promoter operably linked to the nucleic acid sequence

encoding at least a portion of the target gene, or a cofactor of the target gene, a mutant version of at least a portion of the target gene, or in some embodiments, an inhibitor of the target gene. Alternatively, the construct can be promoterless. In cases in which the construct is not designed to incorporate into the cellular DNA/genome, the vector typically contains at least one promoter element. In addition to the nucleic acid sequences, the expression vector can contain selectable marker sequences. The expression constructs can further contain sites for transcription initiation, termination, and/or ribosome binding sites. The identified constructs can be inserted into and can be expressed in any prokaryotic or eukaryotic cell, including, but not limited to bacterial cells, such as P. fluorescens or E. coli, yeast cells, mammalian cells, such as CHO cells, or plant cells.

The construct can be prepared in accordance with processes known in the art. Various fragments can be assembled, introduced into appropriate vectors, cloned, analyzed and then manipulated further until the desired construct has been achieved. Various modifications can be made to the sequence, to allow for restriction analysis, excision, identification of probes, etc. Silent mutations can be introduced, as desired. At various stages, restriction analysis, sequencing, amplification with the polymerase chain reaction, primer repair, *in vitro* mutagenesis, etc. can be employed. Processes for the incorporation of antibiotic resistance genes and negative selection factors will be familiar to those of ordinary skill in the art (see, e.g., WO 99/15650; U.S. Pat. No. 6,080,576; U.S. Pat. No. 6,136,566; Niwa, et al., J. Biochem. 113:343-349 (1993); and Yoshida, et al., Transgenic Research, 4:277-287 (1995)).

The construct can be prepared using a bacterial vector, including a prokaryotic replication system, e.g. an origin recognizable by a prokaryotic cell such as P. fluorescens or E. coli. A marker, the same as or different from the marker to be used for insertion, can be employed, which can be removed prior to introduction into the host cell. Once the vector containing the construct has been completed, it can be further manipulated, such as by deletion of certain sequences, linearization, or by introducing mutations, deletions or other sequences into the homologous sequence. In one embodiment, the target gene construct and the heterologous protein construct are part of the same expression vector, and may or may not be under the control of the same promoter element. In another embodiment, they are on separate expression vectors. After final manipulation, the construct can be introduced into the cell.

Cell growth conditions

The cell growth conditions for the host cells described herein include that which facilitates expression of the protein of interest in at least one strain in the array (or at least a proportion of cells thereof), and/or that which facilitates fermentation of the expressed protein of interest. As used herein, the term "fermentation" includes both embodiments in which literal fermentation is employed and embodiments in which other, non-fermentative culture modes are employed. Growth, maintenance, and/or fermentation of the populations of host cells in the array may be performed at any scale. However, where multiple populations of host cells are screened simultaneously, the scale will be limited by the number of different populations and the capacity to grow and test multiple populations of host cells. In one embodiment, the fermentation medium may be selected from among rich media, minimal media, and mineral salts media. In another embodiment either a minimal medium or a mineral salts medium is selected. In still another embodiment, a minimal medium is selected. In yet another embodiment, a mineral salts medium is selected.

Mineral salts media consists of mineral salts and a carbon source such as, e.g., glucose, sucrose, or glycerol. Examples of mineral salts media include, e.g., M9 medium, Pseudomonas medium (ATCC 179), Davis and Mingioli medium (see, BD Davis & ES Mingioli (1950) in *J. Bact.* 60:17-28). The mineral salts used to make mineral salts media include those selected from among, e.g., potassium phosphates, ammonium sulfate or chloride, magnesium sulfate or chloride, and trace minerals such as calcium chloride, borate, and sulfates of iron, copper, manganese, and zinc. No organic nitrogen source, such as peptone, tryptone, amino acids, or a yeast extract, is included in a mineral salts medium. Instead, an inorganic nitrogen source is used and this may be selected from among, e.g., ammonium salts, aqueous ammonia, and gaseous ammonia. A preferred mineral salts medium will contain glucose as the carbon source. In comparison to mineral salts media, minimal media can also contain mineral salts and a carbon source, but can be supplemented with, e.g., low levels of amino acids, vitamins, peptones, or other ingredients, though these are added at very minimal levels.

In one embodiment, media can be prepared using the components listed in Table 4 below. The components can be added in the following order: first $(NH_4)HPO_4$, KH_2PO_4 and citric acid can be dissolved in approximately 30 liters of distilled water; then a solution of trace elements can be added, followed by the

addition of an antifoam agent, such as Ucolub N 115. Then, after heat sterilization (such as at approximately 121°C.), sterile solutions of glucose $MgSO_4$ and thiamine-HCL can be added. Control of pH at approximately 6.8 can be achieved using aqueous ammonia. Sterile distilled water can then be added to adjust the initial volume to 371 minus the glycerol stock (123 mL). The chemicals are commercially available from various suppliers, such as Merck.

Table 4: Medium composition	
Component	Initial concentration
KH_2PO_4	13.3 g l^{-1}
$(NH_4)_2HPO_4$	4.0 g l^{-1}
Citric Acid	1.7 g l^{-1}
$MgSO_4$-$7H_2O$	1.2g l^{-1}
Trace metal solution	10 ml l^{-1}
Thiamin HCl	4.5 mg l^{-1}
Glucose-H_2O	27.3 g l^{-1}
Antifoam Ucolub N115	0.1 ml l^{-1}
Feeding solution	
$MgSO_4$-$7H_2O$	19.7 g l^{-1}
Glucose-H_2O	770 g l^{-1}
NH_3	23 g
Trace metal solution	
6 g l^{-1} Fe(III) citrate 1.5 g l^{-1} $MnCl_2$-$4H_2O$	
0.8 g l^{-1} $ZmCH_2COOI_2$-$2H_2O$ 0.3 g l^{-1} H_3BO_3	
0.25 g l^{-1} Na_2MoO_4-$2H_2O$ 0.25 g l^{-1} $CoCl_2$ $6H_2O$	
0.15 g l^{-1} $CuCl_2$ $2H_2O$ 0.84 g l^{-1} ethylene	
Dinitrilo-tetracetic acid Na_2 sah $2H_2O$	
(Tritriplex III, Merck)	

In the present invention, growth, culturing, and/or fermentation of the transformed host cells is performed within a temperature range permitting survival of the host cells, preferably a temperature within the range of about 4°C to about 55°C, inclusive. Thus, e.g., the terms "growth" (and "grow," "growing"), "culturing" (and "culture"), and "fermentation" (and "ferment," "fermenting"), as used herein in regard to the host cells of the present invention, inherently means "growth," "culturing," and "fermentation," within a temperature range of about 4°C to about 55°C, inclusive. In addition, "growth" is used to indicate both biological states of active cell division and/or enlargement, as well as biological states in which a non-dividing and/or non-enlarging cell is being metabolically sustained, the latter use of the term "growth" being synonymous with the term "maintenance."

The host cells of the array should be grown and maintained at a suitable temperature for normal growth of that cell type. Such normal growth temperatures may be readily selected based on the known growth requirements of the selected host cell. Preferably, during the establishment of the culture and particularly during course of the screening, the cell culture is incubated in a controlled CO_2/N_2 humidity suitable for growth of the selected cells before and after transformation with the heterologous protein or polypeptide of interest. The humidity of the incubation is controlled to minimize evaporation from the culture vessel, and permit the use of smaller volumes. Alternatively, or in addition to controlling humidity, the vessels may be covered with lids in order to minimize evaporation. Selection of the incubation temperature depends primarily upon the identity of the host cells utilized. Selection of the percent humidity to control evaporation is based upon the selected volume of the vessel and concentration and volume of the cell culture in the vessel, as well as upon the incubation temperature. Thus, the humidity may vary from about 10% to about 80%. It should be understood that selection of a suitable conditions is well within the skill of the art.

Screening

The strain array described herein can be screened for the optimal host cell population in which to express a heterologous protein of interest. The optimal host cell population can be identified or selected based on the quantity, quality, and/or location of the expressed protein of interest. In one embodiment, the optimal host cell

population is one that results in an increased yield of the protein or polypeptide of interest within the host cell compared to other populations of phenotypically distinct host cells in the array.

The increased production alternatively can be an increased level of properly processed protein or polypeptide per gram of protein produced, or per gram of host protein. The increased production can also be an increased level of recoverable protein or polypeptide produced per gram of heterologous protein or per gram of host cell protein. The increased production can also be any combination of an increased level of total protein, increased level of properly processed or properly folded protein, or increased level of active or soluble protein. In this embodiment, the term "increased" or "improved" is relative to the level of protein or polypeptide that is produced, properly processed, soluble, and/or recoverable when the protein or polypeptide of interest is expressed in one or more other populations of host cells in the array. The increased production may optimize the efficiency of the cell or organism by for example, decreasing the energy expenditure, increasing the use of available resources, or decreasing the requirements for growth supplements in growth media. The increased production may also be the result of a decrease in proteolyic degradation of the expressed protein.

In one embodiment, at least one strain in the array produces at least 0.1 mg/ml correctly processed protein. A correctly processed protein has an amino terminus of the native protein. In another embodiment, at least one strain produces 0.1 to 10 mg/ml correctly processed protein in the cell, including at least about 0.2, about 0.3, about 0.4, about 0.5, about 0.6, about 0.7, about 0.8, about 0.9 or at least about 1.0 mg/ml correctly processed protein. In another embodiment, the total correctly processed protein or polypeptide of interest produced by at least one strain in the array is at least 1.0 mg/ml, at least about 2 mg/ml, at least about 3 mg/ml, about 4 mg/ml, about 5 mg/ml, about 6 mg/ml, about 7 mg/ml, about 8 mg/ml, about 10 mg/ml, about 15 mg/ml, about 20 mg/ml, about 25 mg/ml, about 30 mg/ml, about 35 mg/ml, about 40 mg/ml, about 45 mg/ml, at least about 50 mg/ml, or greater. In some embodiments, the amount of correctly processed protein produced is at least about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, about 96%, about 97%, about 98%, at least about 99%, or more of total heterologous protein in a correctly processed form.

An improved expression of a protein or polypeptide of interest can also refer to an increase in the solubility of the protein. The protein or polypeptide of interest can be produced and recovered from the cytoplasm, periplasm or extracellular medium of the host cell. The protein or polypeptide can be insoluble or soluble. The protein or polypeptide can include one or more targeting (e.g., signal or leader) sequences or sequences to assist purification, as discussed *supra*.

The term "soluble" as used herein means that the protein is not precipitated by centrifugation at between approximately 5,000 and 20,000 *x* gravity when spun for 10-30 minutes in a buffer under physiological conditions. Soluble proteins are not part of an inclusion body or other precipitated mass. Similarly, "insoluble" means that the protein or polypeptide can be precipitated by centrifugation at between 5,000 and 20,000 *x* gravity when spun for 10-30 minutes in a buffer under physiological conditions. Insoluble proteins or polypeptides can be part of an inclusion body or other precipitated mass. The term "inclusion body" is meant to include any intracellular body contained within a cell wherein an aggregate of proteins or polypeptides has been sequestered.

In another embodiment, the optimal host cell population produces an increased amount of the protein of interest that is transported to the periplasm or secreted into the extracellular space of the host cell. In one embodiment, at least one strain in the array produces at least 0.1 mg/ml protein in the periplasmic compartment. In another embodiment, at least one strain produces 0.1 to 10 mg/ml periplasmic protein in the cell, or at least about 0.2, about 0.3, about 0.4, about 0.5, about 0.6, about 0.7, about 0.8, about 0.9 or at least about 1.0 mg/ml periplasmic protein. In one embodiment, the total protein or polypeptide of interest produced by at least one strain in the array is at least 1.0 mg/ml, at least about 2 mg/ml, at least about 3 mg/ml, about 4 mg/ml, about 5 mg/ml, about 6 mg/ml, about 7 mg/ml, about 8 mg/ml, about 10 mg/ml, about 15 mg/ml, about 20 mg/ml, at least about 25 mg/ml, or greater. In some embodiments, the amount of periplasmic protein produced is at least about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99%, or more of total protein or polypeptide of interest produced.

At least one strain in the array of the invention can also lead to increased yield of the protein or polypeptide of interest. In one embodiment, at least one strain produces a protein or polypeptide of interest as at least about 5%, at least about 10%,

about 15%, about 20%, about 25%, about 30%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, or greater of total cell protein (tcp). "Percent total cell protein" is the amount of protein or polypeptide in the host cell as a percentage of aggregate cellular protein. Methods for the determination of the percent total cell protein are well known in the art.

In a particular embodiment, at least one host cell population in the array can have a heterologous protein production level of at least 1% tcp and a cell density of at least 40 mg/ml, when grown (i.e. within a temperature range of about 4°C to about 55°C, including about 10°C, about 15°C, about 20°C, about 25°C, about 30°C, about 35°C, about 40°C, about 45°C, and about 50°C) in a mineral salts medium. In a particularly preferred embodiment, the expression system will have a protein or polypeptide expression level of at least 5% tcp and a cell density of at least 40 g/L, when grown (i.e. within a temperature range of about 4°C to about 55°C, inclusive) in a mineral salts medium.

In practice, heterologous proteins targeted to the periplasm are often found in the broth (see European Patent No. EP 0 288 451), possibly because of damage to or an increase in the fluidity of the outer cell membrane. The rate of this "passive" secretion may be increased by using a variety of mechanisms that permeabilize the outer cell membrane, including: colicin (Miksch *et al.* (1997) *Arch. Microbiol.* 167: 143-150); growth rate (Shokri *et al.* (2002) *App Miocrobiol Biotechnol* 58:386-392); TolIII overexpression (Wan and Baneyx (1998) *Protein Expression Purif.* 14: 13-22); bacteriocin release protein (Hsiung *et al.* (1989) *Bio/Technology* 7: 267-71), colicin A lysis protein (Lloubes *et al.* (1993) *Biochimie* 75: 451-8) mutants that leak periplasmic proteins (Furlong and Sundstrom (1989) Developments in Indus. *Microbio.* 30: 141-8); fusion partners (Jeong and Lee (2002) *Appl. Environ. Microbio*. 68: 4979-4985); or, recovery by osmotic shock (Taguchi *et al.* (1990) *Biochimica Biophysica Acta* 1049: 278-85). Transport of engineered proteins to the periplasmic space with subsequent localization in the broth has been used to produce properly folded and active proteins in *E. coli* (Wan and Baneyx (1998) *Protein Expression Purif.* 14: 13-22; Simmons *et al.* (2002) *J. Immun. Meth*.263: 133-147; Lundell *et al.* (1990) *J. Indust. Microbio*. 5: 215-27).

The method may also include the step of purifying the protein or polypeptide of interest from the periplasm or from extracellular media. The heterologous protein or polypeptide can be expressed in a manner in which it is linked to a tag protein and the "tagged" protein can be purified from the cell or extracellular media.

In some embodiments, the protein or polypeptide of interest can also be produced by at least one strain in the array in an active form. The term "active" means the presence of biological activity, wherein the biological activity is comparable or substantially corresponds to the biological activity of a corresponding native protein or polypeptide. In the context of proteins this typically means that a polynucleotide or polypeptide comprises a biological function or effect that has at least about 20%, about 50%, preferably at least about 60-80%, and most preferably at least about 90-95% activity compared to the corresponding native protein or polypeptide using standard parameters. However, in some embodiments, it may be desirable to produce a polypeptide that has altered or improved activity compared to the native protein (e.g, one that has altered or improved immunoreactivity, substrate specificity, etc). An altered or improved polypeptide may result from a particular conformation created by one or more of the host cell populations of the array.

The determination of protein or polypeptide activity can be performed utilizing corresponding standard, targeted comparative biological assays for particular proteins or polypeptides which can be used to assess biological activity.

The recovery of active protein or polypeptide of interest may also be improved in the optimal host strain compared to one or more other strains in the array of the invention. Active proteins can have a specific activity of at least about 20%, at least about 30%, at least about 40%, about 50%, about 60%, at least about 70%, about 80%, about 90%, or at least about 95% that of the native protein or polypeptide from which the sequence is derived. Further, the substrate specificity (k_{cat}/K_m) is optionally substantially similar to the native protein or polypeptide. Typically, k_{cat}/K_m will be at least about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, at least about 90%, at least about 95%, or greater. Methods of assaying and quantifying measures of protein and polypeptide activity and substrate specificity (k_{cat}/K_m), are well known to those of skill in the art.

Measurement of protein activity

The activity of the heterologously-expressed protein or polypeptide of interest can be compared with a previously established native protein or polypeptide standard activity. Alternatively, the activity of the protein or polypeptide of interest can be determined in a simultaneous, or substantially simultaneous, comparative assay with the native protein or polypeptide. For example, *in vitro* assays can be used to determine any detectable interaction between a protein or polypeptide of interest and a target, e.g. between an expressed enzyme and substrate, between expressed hormone and hormone receptor, between expressed antibody and antigen, etc. Such detection can include the measurement of calorimetric changes, proliferation changes, cell death, cell repelling, changes in radioactivity, changes in solubility, changes in molecular weight as measured by gel electrophoresis and/or gel exclusion methods, phosphorylation abilities, antibody specificity assays such as ELISA assays, etc. In addition, in vivo assays include, but are not limited to, assays to detect physiological effects of the heterologously expressed protein or polypeptide in comparison to physiological effects of the native protein or polypeptide, e.g. weight gain, change in electrolyte balance, change in blood clotting time, changes in clot dissolution and the induction of antigenic response. Generally, any in vitro or in vivo assay can be used to determine the active nature of the protein or polypeptide of interest that allows for a comparative analysis to the native protein or polypeptide so long as such activity is assayable. Alternatively, the proteins or polypeptides produced in at least one strain in the array of the present invention can be assayed for the ability to stimulate or inhibit interaction between the protein or polypeptide and a molecule that normally interacts with the protein or polypeptide, e.g. a substrate or a component of a signal pathway with which the native protein normally interacts. Such assays can typically include the steps of combining the protein with a substrate molecule under conditions that allow the protein or polypeptide to interact with the target molecule, and detect the biochemical consequence of the interaction with the protein and the target molecule.

Assays that can be utilized to determine protein or polypeptide activity are described, for example, in Ralph, P. J., *et al.* (1984) *J. Immunol.* 132:1858 or Saiki *et al.* (1981) *J. Immunol.* 127:1044, Steward, W. E. II (1980) *The Interferon Systems*. Springer-Verlag, Vienna and New York, Broxmeyer, H. E., *et al.* (1982) *Blood* 60:595, *Molecular Cloning: A Laboratory Manua*", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and *Methods*

in Enzymology: Guide to Molecular Cloning Techniques, Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987, A K Patra *et al.*, *Protein Expr Purif*, 18(2): p/182-92 (2000), Kodama *et al.*, *J. Biochem*. 99: 1465-1472 (1986); Stewart *et al.*, *Proc. Nat'l Acad. Sci*. USA 90: 5209-5213 (1993); (Lombillo *et al.*, *J. Cell Biol*. 128:107-115 (1995); (Vale *et al.*, *Cell* 42:39-50 (1985). Activity can be compared between samples of heterologously expressed protein derived from one or more of the other host cell populations in the array, or can be compared to the activity of a native protein, or both. Activity measurements can be performed on isolated protein, or can be performed *in vitro* in the host cell.

In another embodiment, protein production and/or activity may be monitored directly in the culture by fluorescence or spectroscopic measurements on, for example, a conventional microscope, luminometer, or plate reader. Where the protein of interest is an enzyme whose substrate is known, the substrate can be added to the culture media wherein a fluorescent signal is emitted when the substrate is converted by the enzyme into a product. In one embodiment, the expression construct encoding the heterologous protein or polypeptide of interest further encodes a reported protein. By "reporter protein" is meant a protein that by its presence in or on a cell or when secreted in the media allows the cell to be distinguished from a cell that does not contain the reporter protein. Production of the heterologous protein of interest results in a detectable change in the host cell population. The reporter molecule can be firefly luciferase and GFP or any other fluorescence molecule, as well as beta-galactosidase gene (beta.gal) and chloramphenicol and acetyltransferase gene (CAT). Assays for expression produced in conjunction with each of these reporter gene elements are well-known to those skilled in the art.

The reporter gene can encode a detectable protein or an indirectly detectable protein, or the reporter gene can be a survival gene. In a preferred embodiment, the reporter protein is a detectable protein. A "detectable protein" or "detection protein" (encoded by a detectable or detection gene) is a protein that can be used as a direct label; that is, the protein is detectable (and preferably, a cell comprising the detectable protein is detectable) without further manipulation. Thus, in this embodiment, the protein product of the reporter gene itself can serve to distinguish cells that are expressing the detectable gene. In this embodiment, suitable detectable genes include those encoding autofluorescent proteins.

As is known in the art, there are a variety of autofluorescent proteins known; these generally are based on the green fluorescent protein (GFP) from Aequorea and variants thereof; including, but not limited to, GFP, (Chalfie, *et al.* (1994) *Science* 263(5148):802-805); enhanced GFP (EGFP; Clontech--Genbank Accession Number U55762)), blue fluorescent protein (BFP; Quantum Biotechnologies, Inc., Montreal, Canada); Stauber (1998) *Biotechniques* 24(3):462-471; Heim and Tsien(1996) *Curr. Biol.* 6:178-182), enhanced yellow fluorescent protein (EYFP; Clontech Laboratories, Inc., Palo Alto, CA) and red fluorescent protein. In addition, there are recent reports of autofluorescent proteins from Renilla and Ptilosarcus species. See WO 92/15673; WO 95/07463; WO 98/14605; WO 98/26277; WO 99/49019; U.S. Pat. No. 5,292,658; U.S. Pat. No. 5,418,155; U.S. Pat. No. 5,683,888; U.S. Pat. No. 5,741,668; U.S. Pat. No. 5,777,079; U.S. Pat. No. 5,804,387; U.S. Pat. No. 5,874,304; U.S. Pat. No. 5,876,995; and U.S. Pat. No. 5,925,558; all of which are expressly incorporated herein by reference.

Isolation of Protein or Polypeptide of Interest

To measure the yield, solubility, conformation, and/or activity of the protein of interest, it may be desirable to isolate the protein from one or more strains in the array. The isolation may be a crude, semi-crude, or pure isolation, depending on the requirements of the assay used to make the appropriate measurements. The protein may be produced in the cytoplasm, targeted to the periplasm, or may be secreted into the culture or fermentation media. To release proteins targeted to the periplasm, treatments involving chemicals such as chloroform (Ames *et al.* (1984) *J. Bacteriol.*, 160: 1181-1183), guanidine-HCl, and Triton X-100 (Naglak and Wang (1990) *Enzyme Microb. Technol.*, 12: 603-611) have been used. However, these chemicals are not inert and may have detrimental effects on many heterologous protein products or subsequent purification procedures. Glycine treatment of *E. coli* cells, causing permeabilization of the outer membrane, has also been reported to release the periplasmic contents (Ariga *et al.* (1989) J. Ferm. Bioeng., 68: 243-246). The most widely used methods of periplasmic release of heterologous protein are osmotic shock (Nosal and Heppel (1966) *J. Biol. Chem.*, 241: 3055-3062; Neu and Heppel (1965) *J. Biol. Chem.*, 240: 3685-3692), hen eggwhite (HEW)-lysozyme/ethylenediamine tetraacetic acid (EDTA) treatment (Neu and Heppel (1964) *J. Biol. Chem.*, 239: 3893-3900; Witholt *et al.* (1976) *Biochim. Biophys. Acta*, 443: 534-544; Pierce *et al.* (1995)

ICheme Research. Event, 2: 995-997), and combined HEW-lysozyme/osmotic shock treatment (French *et al.* (1996) *Enzyme and Microb. Tech.*, 19: 332-338). The French method involves resuspension of the cells in a fractionation buffer followed by recovery of the periplasmic fraction, where osmotic shock immediately follows lysozyme treatment.

Typically, these procedures include an initial disruption in osmotically-stabilizing medium followed by selective release in non-stabilizing medium. The composition of these media (pH, protective agent) and the disruption methods used (chloroform, HEW-lysozyme, EDTA, sonication) vary among specific procedures reported. A variation on the HEW-lysozyme/EDTA treatment using a dipolar ionic detergent in place of EDTA is discussed by Stabel *et al.* (1994) *Veterinary Microbiol.*, 38: 307-314. For a general review of use of intracellular lytic enzyme systems to disrupt *E. coli*, see Dabora and Cooney (1990) in *Advances in Biochemical Engineering/Biotechnology*, Vol. 43, A. Fiechter, ed. (Springer-Verlag: Berlin), pp. 11-30.

Conventional methods for the recovery of proteins or polypeptides of interest from the cytoplasm, as soluble protein or refractile particles, involved disintegration of the bacterial cell by mechanical breakage. Mechanical disruption typically involves the generation of local cavitation in a liquid suspension, rapid agitation with rigid beads, sonication, or grinding of cell suspension (*Bacterial Cell Surface Techniques*, Hancock and Poxton (John Wiley & Sons Ltd, 1988), Chapter 3, p. 55).

HEW-lysozyme acts biochemically to hydrolyze the peptidoglycan backbone of the cell wall. The method was first developed by Zinder and Arndt (1956) *Proc. Natl. Acad. Sci. USA*, 42: 586-590, who treated *E. coli* with egg albumin (which contains HEW-lysozyme) to produce rounded cellular spheres later known as spheroplasts. These structures retained some cell-wall components but had large surface areas in which the cytoplasmic membrane was exposed. U.S. Pat. No. 5,169,772 discloses a method for purifying heparinase from bacteria comprising disrupting the envelope of the bacteria in an osmotically-stabilized medium, e.g., 20% sucrose solution using, e.g., EDTA, lysozyme, or an organic compound, releasing the non-heparinase-like proteins from the periplasmic space of the disrupted bacteria by exposing the bacteria to a low-ionic-strength buffer, and releasing the heparinase-like proteins by exposing the low-ionic-strength-washed bacteria to a buffered salt solution.

Many different modifications of these methods have been used on a wide range of expression systems with varying degrees of success (Joseph-Liazun *et al.* (1990) *Gene*, 86: 291-295; Carter *et al.* (1992) *Bio/Technology*, 10: 163-167). Efforts to induce recombinant cell culture to produce lysozyme have been reported. EP 0 155 189 discloses a means for inducing a recombinant cell culture to produce lysozymes, which would ordinarily be expected to kill such host cells by means of destroying or lysing the cell wall structure.

U.S. Pat. No. 4,595,658 discloses a method for facilitating externalization of proteins transported to the periplasmic space of bacteria. This method allows selective isolation of proteins that locate in the periplasm without the need for lysozyme treatment, mechanical grinding, or osmotic shock treatment of cells. U.S. Pat. No. 4,637,980 discloses producing a bacterial product by transforming a temperature-sensitive lysogen with a DNA molecule that codes, directly or indirectly, for the product, culturing the transformant under permissive conditions to express the gene product intracellularly, and externalizing the product by raising the temperature to induce phage-encoded functions. Asami *et al.* (1997) *J. Ferment. and Bioeng.*, 83: 511-516 discloses synchronized disruption of *E. coli* cells by T4 phage infection, and Tanji *et al.* (1998) *J. Ferment. and Bioeng.*, 85: 74-78 discloses controlled expression of lysis genes encoded in T4 phage for the gentle disruption of *E. coli* cells.

Upon cell lysis, genomic DNA leaks out of the cytoplasm into the medium and results in significant increase in fluid viscosity that can impede the sedimentation of solids in a centrifugal field. In the absence of shear forces such as those exerted during mechanical disruption to break down the DNA polymers, the slower sedimentation rate of solids through viscous fluid results in poor separation of solids and liquid during centrifugation. Other than mechanical shear force, there exist nucleolytic enzymes that degrade DNA polymer. In *E. coli*, the endogenous gene endA encodes for an endonuclease (molecular weight of the mature protein is approx. 24.5 kD) that is normally secreted to the periplasm and cleaves DNA into oligodeoxyribonucleotides in an endonucleolytic manner. It has been suggested that endA is relatively weakly expressed by *E. coli* (Wackemagel *et al.* (1995) *Gene* 154: 55-59).

If desired, the proteins produced using one or more strains in the array of this invention may be isolated and purified to substantial purity by standard techniques well known in the art, including, but not limited to, ammonium sulfate or ethanol

precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, nickel chromatography, hydroxylapatite chromatography, reverse phase chromatography, lectin chromatography, preparative electrophoresis, detergent solubilization, selective precipitation with such substances as column chromatography, immunopurification methods, and others. For example, proteins having established molecular adhesion properties can be reversibly fused with a ligand. With the appropriate ligand, the protein can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein is then removed by enzymatic activity. In addition, protein can be purified using immunoaffinity columns or Ni-NTA columns. General techniques are further described in, for example, R. Scopes, *Protein Purification*: Principles and Practice, Springer-Verlag: N.Y. (1982); Deutscher, *Guide to Protein Purification*, Academic Press (1990); U.S. Pat. No. 4,511,503; S. Roe, *Protein Purification Techniques: A Practical Approach* (Practical Approach Series), Oxford Press (2001); D. Bollag, *et al.*, Protein Methods, Wiley-Lisa, Inc. (1996); AK Patra *et al.*, *Protein Expr Purif*, 18(2): p/182-92 (2000); and R. Mukhija, *et al.*, *Gene* 165(2): p. 303-6 (1995). See also, for example, Ausubel, *et al.* (1987 and periodic supplements); Deutscher (1990) "Guide to Protein Purification," *Methods in Enzymology* vol. 182, and other volumes in this series; Coligan, *et al.* (1996 and periodic Supplements) *Current Protocols in Protein Science* Wiley/Greene, NY; and manufacturer's literature on use of protein purification products, e.g., Pharmacia, Piscataway, N.J., or Bio-Rad, Richmond, Calif. Combination with recombinant techniques allow fusion to appropriate segments, e.g., to a FLAG sequence or an equivalent which can be fused via a protease-removable sequence. See also, for example., Hochuli (1989) *Chemische Industrie* 12:69-70; Hochuli (1990) "Purification of Recombinant Proteins with Metal Chelate Absorbent" in Setlow (ed.) *Genetic Engineering, Principle and Methods* 12:87-98, Plenum Press, NY; and Crowe, *et al.* (1992) QIAexpress: The High Level Expression & Protein Purification System QUIAGEN, Inc., Chatsworth, Calif.

Detection of the expressed protein is achieved by methods known in the art and include, for example, radioimmunoassays, Western blotting techniques or immunoprecipitation.

Certain proteins expressed by the strains in the array of this invention may form insoluble aggregates ("inclusion bodies"). Several protocols are suitable for purification of proteins from inclusion bodies. For example, purification of inclusion bodies typically involves the extraction, separation and/or purification of inclusion bodies by disruption of the host cells, e.g., by incubation in a buffer of 50 mM TRIS/HCL pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mM ATP, and 1 mM PMSF. The cell suspension is typically lysed using 2-3 passages through a French Press. The cell suspension can also be homogenized using a Polytron (Brinkman Instruments) or sonicated on ice. Alternate methods of lysing bacteria are apparent to those of skill in the art (see, e.g., Sambrook *et al.*, supra; Ausubel *et al.*, supra).

If necessary, the inclusion bodies can be solubilized, and the lysed cell suspension typically can be centrifuged to remove unwanted insoluble matter. Proteins that formed the inclusion bodies may be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of immunologically and/or biologically active protein. Other suitable buffers are known to those skilled in the art.

The heterologously-expressed proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art. For example, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the protein or polypeptide of interest. One such example can be ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol includes adding saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This concentration will precipitate the most hydrophobic of proteins. The precipitate is then discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in

buffer and the excess salt removed if necessary, either through dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

The molecular weight of a protein or polypeptide of interest can be used to isolated it from proteins of greater and lesser size using ultrafiltration through membranes of different pore size (for example, Amicon or Millipore membranes). As a first step, the protein mixture can be ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration can then be ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The protein or polypeptide of interest will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

The expressed proteins or polypeptides of interest can also be separated from other proteins on the basis of its size, net surface charge, hydrophobicity, and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art. It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g., Pharmacia Biotech).

Renaturation and Refolding

Where heterologously expressed protein is produced in a denatured form, insoluble protein can be renatured or refolded to generate secondary and tertiary protein structure conformation. Protein refolding steps can be used, as necessary, in completing configuration of the heterologous product. Refolding and renaturation can be accomplished using an agent that is known in the art to promote dissociation/association of proteins. For example, the protein can be incubated with dithiothreitol followed by incubation with oxidized glutathione disodium salt followed by incubation with a buffer containing a refolding agent such as urea.

The protein or polypeptide of interest can also be renatured, for example, by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be refolded while immobilized on a column, such as the Ni NTA column by using a linear 6M-1M urea

gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation can be performed over a period of 1.5 hours or more. After renaturation the proteins can be eluted by the addition of 250 mM imidazole. Imidazole can be removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein can be stored at 4°C or frozen at -80°C.

Other methods include, for example, those that may be described in M H Lee *et al., Protein Expr. Purif.*, 25(1): p. 166-73 (2002), W. K. Cho *et al., J. Biotechnology*, 77(2-3): p. 169-78 (2000), Ausubel, *et al.* (1987 and periodic supplements), Deutscher (1990) "Guide to Protein Purification," *Methods in Enzymology* vol. 182, and other volumes in this series, Coligan, *et al.* (1996 and periodic Supplements) *Current Protocols in Protein Science* Wiley/Greene, NY, S. Roe, *Protein Purification Techniques: A Practical Approach* (Practical Approach Series), Oxford Press (2001); D. Bollag, *et al.*, Protein Methods, Wiley-Lisa, Inc. (1996)

Expression vectors

A heterologous protein of interest can be produced in one or more of the host cells disclosed herein by introducing into each strain an expression vector encoding the heterologous protein of interest. In one embodiment, the vector comprises a polynucleotide sequence encoding the protein of interest operably linked to a promoter capable of functioning in the chosen host cell, as well as all other required transcription and translation regulatory elements.

The term "operably linked" refers to any configuration in which the transcriptional and any translational regulatory elements are covalently attached to the encoding sequence in such disposition(s), relative to the coding sequence, that in and by action of the host cell, the regulatory elements can direct the expression of the coding sequence.

The heterologous protein of interest can be expressed from polynucleotides in which the heterologous polypeptide coding sequence is operably linked to transcription and translation regulatory elements to form a functional gene from which the host cell can express the protein or polypeptide. The coding sequence can be a native coding sequence for the heterologous polypeptide, or may be a coding sequence that has been selected, improved, or optimized for use in the selected

expression host cell: for example, by synthesizing the gene to reflect the codon use bias of a host species. In one embodiment of the invention, the host species is a *P. fluorescens*, and the codon bias of *P. fluorescens* is taken into account when designing the polypeptide coding sequence. The gene(s) are constructed within or inserted into one or more vector(s), which can then be transformed into the expression host cell.

Other regulatory elements may be included in a vector (also termed "expression construct"). The vector will typically comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Additional elements include, but are not limited to, for example, transcriptional enhancer sequences, translational enhancer sequences, other promoters, activators, translational start and stop signals, transcription terminators, cistronic regulators, polycistronic regulators, or tag sequences, such as nucleotide sequence "tags" and "tag" polypeptide coding sequences, which facilitates identification, separation, purification, and/or isolation of an expressed polypeptide.

In another embodiment, the expression vector further comprises a tag sequence adjacent to the coding sequence for the protein or polypeptide of interest. In one embodiment, this tag sequence allows for purification of the protein. The tag sequence can be an affinity tag, such as a hexa-histidine affinity tag. In another embodiment, the affinity tag can be a glutathione-S-transferase molecule. The tag can also be a fluorescent molecule, such as YFP or GFP, or analogs of such fluorescent proteins. The tag can also be a portion of an antibody molecule, or a known antigen or ligand for a known binding partner useful for purification.

A protein-encoding gene according to the present invention can include, in addition to the protein coding sequence, the following regulatory elements operably linked thereto: a promoter, a ribosome binding site (RBS), a transcription terminator, translational start and stop signals. Useful RBSs can be obtained from any of the species useful as host cells in expression systems according to the present invention, preferably from the selected host cell. Many specific and a variety of consensus RBSs are known, e.g., those described in and referenced by D. Frishman *et al.*, Gene 234(2):257-65 (8 Jul. 1999); and B. E. Suzek *et al.*, Bioinformatics 17(12):1123-30 (December 2001). In addition, either native or synthetic RBSs may be used, e.g., those described in: EP 0207459 (synthetic RBSs); O. Ikehata *et al.*, Eur. J. Biochem. 181(3):563-70 (1989) (native RBS sequence of AAGGAAG). Further examples of

methods, vectors, and translation and transcription elements, and other elements useful in the present invention are described in, e.g.: U.S. Pat. No. 5,055,294 to Gilroy and U.S. Pat. No. 5,128,130 to Gilroy *et al.*; U.S. Pat. No. 5,281,532 to Rammler *et al.*; U.S. Pat. Nos. 4,695,455 and 4,861,595 to Barnes *et al.*; U.S. Pat. No. 4,755,465 to Gray *et al.*; and U.S. Pat. No. 5,169,760 to Wilcox.

Transcription of the DNA encoding the heterologous protein of interest is increased by inserting an enhancer sequence into the vector or plasmid. Typical enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp in size that act on the promoter to increase its transcription. Examples include various Pseudomonas enhancers.

Generally, the heterologous expression vectors will include origins of replication and selectable markers permitting transformation of the host cell and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding the enzymes such as 3-phosphoglycerate kinase (PGK), acid phosphatase, or heat shock proteins, among others. Where signal sequences are used, the heterologous coding sequence is assembled in appropriate phase with translation initiation and termination sequences, and the signal sequence capable of directing compartmental accumulation or secretion of the translated protein. Optionally the heterologous sequence can encode a fusion enzyme including an N-terminal identification polypeptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed heterologous product. The fusion polypeptide can also comprise one or more target proteins or inhibitors or enhances thereof, as discussed *supra*.

Vectors are known in the art for expressing heterologous proteins in host cells, and any of these may be used for expressing the genes according to the present invention. Such vectors include, e.g., plasmids, cosmids, and phage expression vectors. Examples of useful plasmid vectors include, but are not limited to, the expression plasmids pBBR1MCS, pDSK519, pKT240, pML122, pPS10, RK2, RK6, pRO1600, and RSF1010. Other examples of such useful vectors include those described by, e.g.: N. Hayase, in Appl. Envir. Microbiol. 60(9):3336-42 (September 1994); A. A. Lushnikov *et al.*, in Basic Life Sci. 30:657-62 (1985); S. Graupner & W. Wackemagel, in Biomolec. Eng. 17(1):11-16. (October 2000); H. P. Schweizer, in Curr. Opin. Biotech. 12(5):439-45 (October 2001); M. Bagdasarian & K. N. Timmis,

in Curr. Topics Microbiol. Immunol. 96:47-67 (1982); T. Ishii *et al.*, in FEMS Microbiol. Lett. 116(3):307-13 (Mar. 1, 1994); I. N. Olekhnovich & Y. K. Fomichev, in Gene 140(1):63-65 (Mar. 11, 1994); M. Tsuda & T. Nakazawa, in Gene 136(1-2):257-62 (Dec. 22, 1993); C. Nieto *et al.*, in Gene 87(1):145-49 (Mar. 1, 1990); J. D. Jones & N. Gutterson, in Gene 61(3):299-306 (1987); M. Bagdasarian *et al.*, in Gene 16(1-3):237-47 (December 1981); H. P. Schweizer *et al.*, in Genet. Eng. (NY) 23:69-81 (2001); P. Mukhopadhyay *et al.*, in J. Bact. 172(1):477-80 (January 1990); D. O. Wood *et al.*, in J. Bact. 145(3):1448-51 (March 1981); and R. Holtwick *et al.*, in Microbiology 147(Pt 2):337-44 (February 2001).

Further examples of expression vectors that can be useful in a host cell of the invention include those listed in Table 5 as derived from the indicated replicons.

Table 5. Examples of Useful Expression Vectors

Replicon	Vector(s)
PPS10	PCN39, PCN51
RSF1010	PKT261-3
	PMMB66EH
	PEB8
	PPLGN1
	PMYC1050
RK2/RP1	PRK415
	PJB653
PRO1600	PUCP
	PBSP

The expression plasmid, RSF1010, is described, e.g., by F. Heffron *et al.*, in Proc. Nat'l Acad. Sci. USA 72(9):3623-27 (September 1975), and by K. Nagahari & K. Sakaguchi, in J. Bact. 133(3):1527-29 (March 1978). Plasmid RSF1010 and derivatives thereof are particularly useful vectors in the present invention. Exemplary useful derivatives of RSF1010, which are known in the art, include, e.g., pKT212, pKT214, pKT231 and related plasmids, and pMYC1050 and related plasmids (see, e.g., U.S. Pat. Nos. 5,527,883 and 5,840,554 to Thompson *et al.*), such as, e.g., pMYC1803. Plasmid pMYC1803 is derived from the RSF1010-based plasmid pTJS260 (see U.S. Pat. No. 5,169,760 to Wilcox), which carries a regulated tetracycline resistance marker and the replication and mobilization loci from the RSF1010 plasmid. Other exemplary useful vectors include those described in U.S. Pat. No. 4,680,264 to Puhler *et al.*

In one embodiment, an expression plasmid is used as the expression vector. In another embodiment, RSF1010 or a derivative thereof is used as the expression vector. In still another embodiment, pMYC1050 or a derivative thereof, or pMYC4803 or a derivative thereof, is used as the expression vector.

The plasmid can be maintained in the host cell by inclusion of a selection marker gene in the plasmid. This may be an antibiotic resistance gene(s), where the corresponding antibiotic(s) is added to the fermentation medium, or any other type of selection marker gene known in the art, e.g., a prototrophy-restoring gene where the plasmid is used in a host cell that is auxotrophic for the corresponding trait, e.g., a biocatalytic trait such as an amino acid biosynthesis or a nucleotide biosynthesis trait, or a carbon source utilization trait.

The promoters used in accordance with the present invention may be constitutive promoters or regulated promoters. Common examples of useful regulated promoters include those of the family derived from the lac promoter (i.e. the lacZ promoter), especially the tac and *trc* promoters described in U.S. Pat. No. 4,551,433 to DeBoer, as well as Ptac16, Ptac17, PtacII, PlacUV5, and the T7lac promoter. In one embodiment, the promoter is not derived from the host cell organism. In certain embodiments, the promoter is derived from an *E. coli* organism.

Common examples of non-lac-type promoters useful in expression systems according to the present invention include, e.g., those listed in Table 6.

Table 6. Examples of non-*lac* Promoters	
Promoter	Inducer
P_R	High temperature
P_L	High temperature
Pm	Alkyl- or halo-benzoates
Pu	Alkyl- or halo-toluenes
Psal	Salicylates

See, e.g.: J. Sanchez-Romero & V. De Lorenzo (1999) Manual of Industrial Microbiology and Biotechnology (A. Demain & J. Davies, eds.) pp. 460-74 (ASM Press, Washington, D.C.); H. Schweizer (2001) Current Opinion in Biotechnology, 12:439-445; and R. Slater & R. Williams (2000 Molecular Biology and Biotechnology (J. Walker & R. Rapley, eds.) pp. 125-54 (The Royal Society of Chemistry, Cambridge, UK)). A promoter having the nucleotide sequence of a promoter native to the selected bacterial host cell may also be used to control

expression of the transgene encoding the target polypeptide, e.g, a Pseudomonas anthranilate or benzoate operon promoter (Pant, Pben). Tandem promoters may also be used in which more than one promoter is covalently attached to another, whether the same or different in sequence, e.g., a Pant-Pben tandem promoter (interpromoter hybrid) or a Plac-Plac tandem promoter, or whether derived from the same or different organisms.

Regulated promoters utilize promoter regulatory proteins in order to control transcription of the gene of which the promoter is a part. Where a regulated promoter is used herein, a corresponding promoter regulatory protein will also be part of an expression system according to the present invention. Examples of promoter regulatory proteins include: activator proteins, e.g., *E. coli* catabolite activator protein, MalT protein; AraC family transcriptional activators; repressor proteins, e.g., *E. coli* LacI proteins; and dual-function regulatory proteins, e.g., *E. coli* NagC protein. Many regulated-promoter/promoter-regulatory-protein pairs are known in the art. In one embodiment, the expression construct for the target protein(s) and the heterologous protein of interest are under the control of the same regulatory element.

Promoter regulatory proteins interact with an effector compound, i.e. a compound that reversibly or irreversibly associates with the regulatory protein so as to enable the protein to either release or bind to at least one DNA transcription regulatory region of the gene that is under the control of the promoter, thereby permitting or blocking the action of a transcriptase enzyme in initiating transcription of the gene. Effector compounds are classified as either inducers or co-repressors, and these compounds include native effector compounds and gratuitous inducer compounds. Many regulated-promoter/promoter-regulatory-protein/effector-compound trios are known in the art. Although an effector compound can be used throughout the cell culture or fermentation, in a preferred embodiment in which a regulated promoter is used, after growth of a desired quantity or density of host cell biomass, an appropriate effector compound is added to the culture to directly or indirectly result in expression of the desired gene(s) encoding the protein or polypeptide of interest.

By way of example, where a *lac* family promoter is utilized, a lacI gene can also be present in the system. The lacI gene, which is (normally) a constitutively expressed gene, encodes the *Lac* repressor protein (LacD protein) which binds to the *lac* operator of these promoters. Thus, where a *lac* family promoter is utilized, the

lacI gene can also be included and expressed in the expression system. In the case of the *lac* promoter family members, e.g., the tac promoter, the effector compound is an inducer, preferably a gratuitous inducer such as IPTG (isopropyl-D-1-thiogalactopyranoside, also called "isopropylthiogalactoside").

For expression of a protein or polypeptide of interest, any plant promoter may also be used. A promoter may be a plant RNA polymerase II promoter. Elements included in plant promoters can be a TATA box or Goldberg-Hogness box, typically positioned approximately 25 to 35 basepairs upstream (5') of the transcription initiation site, and the CCAAT box, located between 70 and 100 basepairs upstream. In plants, the CCAAT box may have a different consensus sequence than the functionally analogous sequence of mammalian promoters (Messing *et al.* (1983) In: *Genetic Engineering of Plants*, Kosuge *et al.*, eds., pp. 211-227). In addition, virtually all promoters include additional upstream activating sequences or enhancers (Benoist and Chambon (1981) *Nature* 290:304-310; Gruss *et al.* (1981) *Proc. Nat. Acad. Sci.* 78:943-947; and Khoury and Gruss (1983) *Cell* 27:313-314) extending from around -100 bp to -1,000 bp or more upstream of the transcription initiation site.

Expression Systems

It may be desirable to target the protein or polypeptide of interest to the periplasm of one or more of the populations of host cells in the array, or into the extracellular space. In one embodiment, the expression vector further comprises a nucleotide sequence encoding a secretion signal sequence polypeptide operably linked to the nucleotide sequence encoding the protein or polypeptide of interest. In some embodiments, no modifications are made between the signal sequence and the protein or polypeptide of interest. However, in certain embodiments, additional cleavage signals are incorporated to promote proper processing of the amino terminal of the polypeptide.

The vector can have any of the characteristics described above. In one embodiment, the vector comprising the coding sequence for the protein or polypeptide of interest further comprises a signal sequence, e.g., a secretion signal sequence.

Therefore, in one embodiment, this isolated polypeptide is a fusion protein of the secretion signal and a protein or polypeptide of interest. However, the secretion

signal can also be cleaved from the protein when the protein is targeted to the periplasm. In one embodiment, the linkage between the Sec system secretion signal and the protein or polypeptide is modified to increase cleavage of the secretion signal.

The CHAMPION™ pET expression system provides a high level of protein production. Expression is induced from the strong T7lac promoter. This system takes advantage of the high activity and specificity of the bacteriophage T7 RNA polymerase for high level transcription of the gene of interest. The lac operator located in the promoter region provides tighter regulation than traditional T7-based vectors, improving plasmid stability and cell viability (Studier and Moffatt (1986) *J Molecular Biology* 189(1): 113-30; Rosenberg, *et al.* (1987) *Gene* 56(1): 125-35). The T7 expression system uses the T7 promoter and T7 RNA polymerase (T7 RNAP) for high-level transcription of the gene of interest. High-level expression is achieved in T7 expression systems because the T7 RNAP is more processive than native *E. coli* RNAP and is dedicated to the transcription of the gene of interest. Expression of the identified gene is induced by providing a source of T7 RNAP in the host cell. This is accomplished by using a BL21 *E. coli* host containing a chromosomal copy of the T7 RNAP gene. The T7 RNAP gene is under the control of the lacUV5 promoter which can be induced by IPTG. T7 RNAP is expressed upon induction and transcribes the gene of interest.

The pBAD expression system allows tightly controlled, titratable expression of protein or polypeptide of interest through the presence of specific carbon sources such as glucose, glycerol and arabinose (Guzman, *et al.* (1995) J Bacteriology 177(14): 4121-30). The pBAD vectors are uniquely designed to give precise control over expression levels. Heterologous gene expression from the pBAD vectors is initiated at the araBAD promoter. The promoter is both positively and negatively regulated by the product of the araC gene. AraC is a transcriptional regulator that forms a complex with L-arabinose. In the absence of L-arabinose, the AraC dimer blocks transcription. For maximum transcriptional activation two events are required: (i) L-arabinose binds to AraC allowing transcription to begin, and, (ii) The cAMP activator protein (CAP)-cAMP complex binds to the DNA and stimulates binding of AraC to the correct location of the promoter region.

The *trc* expression system allows high-level, regulated expression in *E. coli* from the *trc* promoter. The *trc* expression vectors have been optimized for expression

of eukaryotic genes in *E. coli*. The *trc* promoter is a strong hybrid promoter derived from the tryptophane (*trp*) and lactose (*lac*) promoters. It is regulated by the lacO operator and the product of the lacIQ gene (Brosius, J. (1984) Gene 27(2): 161-72).

Transformation of the host cells with the vector(s) disclosed herein may be performed using any transformation methodology known in the art, and the bacterial host cells may be transformed as intact cells or as protoplasts (i.e. including cytoplasts). Exemplary transformation methodologies include poration methodologies, e.g., electroporation, protoplast fusion, bacterial conjugation, and divalent cation treatment, e.g., calcium chloride treatment or CaCl/Mg2+ treatment, or other well known methods in the art. See, e.g., Morrison, *J. Bact.*, 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology*, 101:347-362 (Wu *et al.*, eds, 1983), Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

Proteins of interest

The methods and compositions of the present invention are useful for identifying a *P. fluorescens* strain that is optimal for producing high levels of a properly processed protein or polypeptide of interest. The arrays are useful for screening for production of a protein or polypeptide of interest of any species and of any size. However, in certain embodiments, the protein or polypeptide of interest is a therapeutically useful protein or polypeptide. In some embodiments, the protein can be a mammalian protein, for example a human protein, and can be, for example, a growth factor, a cytokine, a chemokine or a blood protein. The protein or polypeptide of interest can be processed in a similar manner to the native protein or polypeptide. In certain embodiments, the protein or polypeptide of interest is less than 100 kD, less than 50 kD, or less than 30 kD in size. In certain embodiments, the protein or polypeptide of interest is a polypeptide of at least about 5, 10, 15, 20, 30, 40, 50 or 100 or more amino acids.

The coding sequence for the protein or polypeptide of interest can be a native coding sequence for the polypeptide, if available, but will more preferably be a coding sequence that has been selected, improved, or optimized for use in an expressible form in the strains of the array: for example, by optimizing the gene to reflect the codon use bias of a Pseudomonas species such as *P. fluorescens* or other suitable

organism. For gene optimization, one or more rare codons may be removed to avoid ribosomal stalling and minimize amino acid misincorporation. One or more gene-internal ribosome binding sites may also be eliminated to avoid truncated protein products. Long stretches of C and G nucleotides may be removed to avoid RNA polymerase slippage that could result in frame-shifts. Strong gene-internal stem-loop structures, especially the ones covering the ribosome binding site, may also be eliminated.

In other embodiments, the protein when produced also includes an additional targeting sequence, for example a sequence that targets the protein to the periplasm or the extracellular medium. In one embodiment, the additional targeting sequence is operably linked to the carboxy-terminus of the protein. In another embodiment, the protein includes a secretion signal for an autotransporter, a two partner secretion system, a main terminal branch system or a fimbrial usher porin.

The gene(s) that result are constructed within or are inserted into one or more vectors, and then transformed into each of the host cell populations in the array. Nucleic acid or a polynucleotide said to be provided in an "expressible form" means nucleic acid or a polynucleotide that contains at least one gene that can be expressed by the one or more of the host cell populations of the invention.

Extensive sequence information required for molecular genetics and genetic engineering techniques is widely publicly available. Access to complete nucleotide sequences of mammalian, as well as human, genes, cDNA sequences, amino acid sequences and genomes can be obtained from GenBank at the website www.ncbi.nlm.nih.gov/Entrez. Additional information can also be obtained from GeneCards, an electronic encyclopedia integrating information about genes and their products and biomedical applications from the Weizmann Institute of Science Genome and Bioinformatics (bioinformatics.weizmann.ac.il/cards), nucleotide sequence information can be also obtained from the EMBL Nucleotide Sequence Database (www.ebi.ac.uk/embl/) or the DNA Databank or Japan (DDBJ, www.ddbi.nig.ac.ii/; additional sites for information on amino acid sequences include Georgetown's protein information resource website (www-nbrf.Reorgetown.edu/pirl) and Swiss-Prot (au.expasy.org/sprot/sprot-top.html).

Examples of proteins that can be expressed in this invention include molecules such as, e.g., renin, a growth hormone, including human growth hormone; bovine growth hormone; growth hormone releasing factor; parathyroid hormone; thyroid

stimulating hormone; lipoproteins; α-1-antitrypsin; insulin A-chain; insulin B-chain; proinsulin; thrombopoietin; follicle stimulating hormone; calcitonin; luteinizing hormone; glucagon; clotting factors such as factor VIIIC, factor IX, tissue factor, and von Willebrands factor; anti-clotting factors such as Protein C; atrial naturietic factor; lung surfactant; a plasminogen activator, such as urokinase or human urine or tissue-type plasminogen activator (t-PA); bombesin; thrombin; hemopoietic growth factor; tumor necrosis factor-alpha and -beta; enkephalinase; a serum albumin such as human serum albumin; mullerian-inhibiting substance; relaxin A-chain; relaxin B-chain; prorelaxin; mouse gonadotropin-associated polypeptide; a microbial protein, such as beta-lactamase; Dnase; inhibin; activin; vascular endothelial growth factor (VEGF); receptors for hormones or growth factors; integrin; protein A or D; rheumatoid factors; a neurotrophic factor such as brain-derived neurotrophic factor (BDNF), neurotrophin-3, -4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6), or a nerve growth factor such as NGF-β; cardiotrophins (cardiac hypertrophy factor) such as cardiotrophin-1 (CT-1); platelet-derived growth factor (PDGF); fibroblast growth factor such as aFGF and bFGF; epidermal growth factor (EGF); transforming growth factor (TGF) such as TGF-alpha and TGF-β, including TGF-β1, TGF-β2, TGF-β3, TGF-β4, or TGF-β5; insulin-like growth factor-I and -II (IGF-I and IGF-II); des(1-3)-IGF-I (brain IGF-I), insulin-like growth factor binding proteins; CD proteins such as CD-3, CD-4, CD-8, and CD-19; erythropoietin; osteoinductive factors; immunotoxins; a bone morphogenetic protein (BMP); an interferon such as interferon-alpha, -beta, and -gamma; colony stimulating factors (CSFs), e.g., M-CSF, GM-CSF, and G-CSF; interleukins (ILs), e.g., IL-1 to IL-10; anti-HER-2 antibody; superoxide dismutase; T-cell receptors; surface membrane proteins; decay accelerating factor; viral antigen such as, for example, a portion of the AIDS envelope; transport proteins; homing receptors; addressins; regulatory proteins; antibodies; and fragments of any of the above-listed polypeptides.

In certain embodiments, the protein or polypeptide can be selected from IL-1, IL-1a, IL-1b, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-12elasti, IL-13, IL-15, IL-16, IL-18, IL-18BPa, IL-23, IL-24, VIP, erythropoietin, GM-CSF, G-CSF, M-CSF, platelet derived growth factor (PDGF), MSF, FLT-3 ligand, EGF, fibroblast growth factor (FGF; e.g., α-FGF (FGF-1), β-FGF (FGF-2), FGF-3, FGF-4, FGF-5, FGF-6, or FGF-7), insulin-like growth factors (e.g., IGF-1,

IGF-2); tumor necrosis factors (e.g., TNF, Lymphotoxin), nerve growth factors (e.g., NGF), vascular endothelial growth factor (VEGF); interferons (e.g., IFN-α, IFN-β, IFN-γ); leukemia inhibitory factor (LIF); ciliary neurotrophic factor (CNTF); oncostatin M; stem cell factor (SCF); transforming growth factors (e.g., TGF-α, TGF-β1, TGF-β2, TGF-β3); TNF superfamily (e.g., LIGHT/TNFSF14, STALL-1/TNFSF13B (BLy5, BAFF, THANK), TNFalpha/TNFSF2 and TWEAK/TNFSF12); or chemokines (BCA-1/BLC-1, BRAK/Kec, CXCL16, CXCR3, ENA-78/LIX, Eotaxin-1, Eotaxin-2/MPIF-2, Exodus-2/SLC, Fractalkine/Neurotactin, GROalpha/MGSA, HCC-1, I-TAC, Lymphotactin/ATAC/SCM, MCP-1/MCAF, MCP-3, MCP-4, MDC/STCP-1/ABCD-1, MIP-1.quadrature., MIP-1.quadrature., MIP-2.quadrature./GRO.quadrature., MIP-3.quadrature./Exodus/LARC, MIP-3/Exodus-3/ELC, MIP-4/PARC/DC-CK1, PF-4, RANTES, SDF1, TARC, TECK, microbial toxins, ADP ribosylating toxins, microbial or viral antigens).

In one embodiment of the present invention, the protein of interest can be a multi-subunit protein or polypeptide. Multisubunit proteins that can be expressed include homomeric and heteromeric proteins. The multisubunit proteins may include two or more subunits that may be the same or different. For example, the protein may be a homomeric protein comprising 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more subunits. The protein also may be a heteromeric protein including 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more subunits. Exemplary multisubunit proteins include: receptors including ion channel receptors; extracellular matrix proteins including chondroitin; collagen; immunomodulators including MHC proteins, full chain antibodies, and antibody fragments; enzymes including RNA polymerases, and DNA polymerases; and membrane proteins.

In another embodiment, the protein of interest can be a blood protein. The blood proteins expressed in this embodiment include but are not limited to carrier proteins, such as albumin, including human and bovine albumin, transferrin, recombinant transferrin half-molecules, haptoglobin, fibrinogen and other coagulation factors, complement components, immunoglobulins, enzyme inhibitors, precursors of substances such as angiotensin and bradykinin, insulin, endothelin, and globulin, including alpha, beta, and gamma-globulin, and other types of proteins, polypeptides, and fragments thereof found primarily in the blood of mammals. The amino acid sequences for numerous blood proteins have been reported (see, S. S. Baldwin (1993)

Comp. Biochem Physiol. 106b:203-218), including the amino acid sequence for human serum albumin (Lawn, L. M., *et al.* (1981) Nucleic Acids Research, 9:6103-6114.) and human serum transferrin (Yang, F. *et al.* (1984) Proc. Natl. Acad. Sci. USA 81:2752-2756).

In another embodiment, the protein of interest can be an enzyme or co-factor. The enzymes and co-factors expressed in this embodiment include but are not limited to aldolases, amine oxidases, amino acid oxidases, aspartases, B12 dependent enzymes, carboxypeptidases, carboxyesterases, carboxylyases, chemotrypsin, CoA requiring enzymes, cyanohydrin synthetases, cystathione synthases, decarboxylases, dehydrogenases, alcohol dehydrogenases, dehydratases, diaphorases, dioxygenases, enoate reductases, epoxide hydrases, fumerases, galactose oxidases, glucose isomerases, glucose oxidases, glycosyltrasferases, methyltransferases, nitrile hydrases, nucleoside phosphorylases, oxidoreductases, oxynitilases, peptidases, glycosyltrasferases, peroxidases, enzymes fused to a therapeutically active polypeptide, tissue plasminogen activator; urokinase, reptilase, streptokinase; catalase, superoxide dismutase; Dnase, amino acid hydrolases (e.g., asparaginase, amidohydrolases); carboxypeptidases; proteases, trypsin, pepsin, chymotrypsin, papain, bromelain, collagenase; neuramimidase; lactase, maltase, sucrase, and arabinofuranosidases.

In another embodiment, the protein of interest can be a single chain, Fab fragment and/or full chain antibody or fragments or portions thereof. A single-chain antibody can include the antigen-binding regions of antibodies on a single stably-folded polypeptide chain. Fab fragments can be a piece of a particular antibody. The Fab fragment can contain the antigen binding site. The Fab fragment can contain 2 chains: a light chain and a heavy chain fragment. These fragments can be linked via a linker or a disulfide bond.

In other embodiments, the protein of interest is a protein that is active at a temperature from about 20 to about 42 °C. In one embodiment, the protein is active at physiological temperatures and is inactivated when heated to high or extreme temperatures, such as temperatures over 65 °C.

In one embodiment, the protein of interest is a protein that is active at a temperature from about 20 to about 42 °C, and/or is inactivated when heated to high or extreme temperatures, such as temperatures over 65 °C; is, or is substantially homologous to, a native protein, such as a native mammalian or human protein and

not expressed from nucleic acids in concatameric form, where the promoter is not a native promoter in to the host cell used in the array but is derived from another organism, such as *E. coli*.

Host cell

In one embodiment the invention provides an array of *P. fluorescens* host cells from which to optimally produce a heterologous protein or peptide of interest. *P. fluorescens* has been demonstrated to be an improved platform for production of a variety of proteins and several efficient secretion signals have been identified from this organism (see, U.S. Application Publication Number 2006/0008877, herein incorporated by reference in its entirety).

The Pseudomonads system offers advantages for commercial expression of polypeptides and enzymes, in comparison with other bacterial expression systems. In particular, *P. fluorescens* has been identified as an advantageous expression system. *P. fluorescens* encompasses a group of common, nonpathogenic saprophytes that colonize soil, water and plant surface environments. Commercial enzymes derived from *P. fluorescens* have been used to reduce environmental contamination, as detergent additives, and for stereoselective hydrolysis. *P. fluorescens* is also used agriculturally to control pathogens. U.S. Patent Number 4,695,462 describes the expression of recombinant bacterial proteins in *P. fluorescens*.

However, it is contemplated that alternate host cells, or even a multiplicity of different host cells, can be used to generate an array comprising a plurality of phenotypically distinct host cells that have been genetically modified to modulate the expression of one or more target genes, as discussed *supra*. The host cell can be any organism in which target genes can be altered. Methods of identifying target genes homologous to those listed in Tables 1 and 2 are known in the art. Further, one of skill in the art would understand how to identify target genes that are native to or useful in a host cell of interest. Many of these proteins are well known in the art. See, for example, U.S. Patent Application Publication No. 2006/0110747).

Host cells can be selected from "Gram-negative Proteobacteria Subgroup 18." "Gram-negative Proteobacteria Subgroup 18" is defined as the group of all subspecies, varieties, strains, and other sub-special units of the species *Pseudomonas fluorescens*, including those belonging, e.g., to the following (with the ATCC or other deposit numbers of exemplary strain(s) shown in parenthesis): *Pseudomonas*

fluorescens biotype A, also called biovar 1 or biovar I (ATCC 13525); *Pseudomonas fluorescens* biotype B, also called biovar 2 or biovar II (ATCC 17816); *Pseudomonas fluorescens* biotype C, also called biovar 3 or biovar III (ATCC 17400); *Pseudomonas fluorescens* biotype F, also called biovar 4 or biovar IV (ATCC 12983); *Pseudomonas fluorescens* biotype G, also called biovar 5 or biovar V (ATCC 17518); *Pseudomonas fluorescens* biovar VI; *Pseudomonas fluorescens* Pf0-1; *Pseudomonas fluorescens* Pf-5 (ATCC BAA-477); *Pseudomonas fluorescens* SBW25; and *Pseudomonas fluorescens* subsp. cellulosa (NCIMB 10462).

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 19." "Gram-negative Proteobacteria Subgroup 19" is defined as the group of all strains of *Pseudomonas fluorescens* biotype A. A particularly preferred strain of this biotype is *P. fluorescens* strain MB101 (see U.S. Pat. No. 5,169,760 to Wilcox), and derivatives thereof. An example of a preferred derivative thereof is *P. fluorescens* strain MB214, constructed by inserting into the MB101 chromosomal *asd* (aspartate dehydrogenase gene) locus, a native *E. coli* PlacI-lacI-lacZYA construct (i.e. in which PlacZ was deleted).

Additional *P. fluorescens* strains that can be used in the present invention include *Pseudomonas fluorescens* Migula and *Pseudomonas fluorescens* Loitokitok, having the following ATCC designations: [NCIB 8286]; NRRL B-1244; NCIB 8865 strain CO1; NCIB 8866 strain CO_2; 1291 [ATCC 17458; IFO 15837; NCIB 8917; LA; NRRL B-1864; pyrrolidine; PW2 [ICMP 3966; NCPPB 967; NRRL B-899]; 13475; NCTC 10038; NRRL B-1603 [6; IFO 15840]; 52-1C; CCEB 488-A [BU 140]; CCEB 553 [EM 15/47]; IAM 1008 [AHH-27]; IAM 1055 [AHH-23]; 1 [IFO 15842]; 12 [ATCC 25323; NIH 11; den Dooren de Jong 216]; 18 [IFO 15833; WRRL P-7]; 93 [TR-10]; 108 [52-22; IFO 15832]; 143 [IFO 15836; PL]; 149 [2-40-40; IFO 15838]; 182 [IFO 3081; PJ 73]; 184 [IFO 15830]; 185 [W2 L-1]; 186 [IFO 15829; PJ 79]; 187 [NCPPB 263]; 188 [NCPPB 316]; 189 [PJ227; 1208]; 191 [IFO 15834; PJ 236; 22/1]; 194 [Klinge R-60; PJ 253]; 196 [PJ 288]; 197 [PJ 290]; 198 [PJ 302]; 201 [PJ 368]; 202 [PJ 372]; 203 [PJ 376]; 204 [IFO 15835; PJ 682]; 205 [PJ 686]; 206 [PJ 692]; 207 [PJ 693]; 208 [PJ 722]; 212. [PJ 832]; 215 [PJ 849]; 216 [PJ 885]; 267 [B-9]; 271 [B-1612]; 401 [C71A; IFO 15831; PJ 187]; NRRL B-3178 [4; IFO. 15841]; KY 8521; 3081; 30-21; [IFO 3081]; N; PYR; PW; D946-B83 [BU 2183; FERM-P 3328]; P-2563 [FERM-P 2894; IFO 13658]; IAM-1126 [43F]; M-1; A506 [A5-06]; A505 [A5-05-1]; A526 [A5-26]; B69; 72; NRRL B-4290; PMW6 [NCIB 11615]; SC 12936; Al

[IFO 15839]; F 1847 [CDC-EB]; F 1848 [CDC 93]; NCIB 10586; P17; F-12; AmMS 257; PRA25; 6133D02; 6519E01; Ni; SC15208; BNL-WVC; NCTC 2583 [NCIB 8194]; H13; 1013 [ATCC 11251; CCEB 295]; IFO 3903; 1062; or Pf-5.

In one embodiment, the host cell can be any cell capable of producing a protein or polypeptide of interest, including a *P. fluorescens* cell as described above. The most commonly used systems to produce proteins or polypeptides of interest include certain bacterial cells, particularly *E. coli*, because of their relatively inexpensive growth requirements and potential capacity to produce protein in large batch cultures. Yeasts are also used to express biologically relevant proteins and polypeptides, particularly for research purposes. Systems include *Saccharomyces cerevisiae* or *Pichia pastoris*. These systems are well characterized, provide generally acceptable levels of total protein production and are comparatively fast and inexpensive. Insect cell expression systems have also emerged as an alternative for expressing recombinant proteins in biologically active form. In some cases, correctly folded proteins that are post-translationally modified can be produced. Mammalian cell expression systems, such as Chinese hamster ovary cells, have also been used for the expression of proteins or polypeptides of interest. On a small scale, these expression systems are often effective. Certain biologics can be derived from proteins, particularly in animal or human health applications. In another embodiment, the host cell is a plant cell, including, but not limited to, a tobacco cell, corn, a cell from an Arabidopsis species, potato or rice cell.

In another embodiment, the host cell can be a prokaryotic cell such as a bacterial cell including, but not limited to, an Escherichia or a Pseudomonas species. Typical bacterial cells are described, for example, in "Biological Diversity: Bacteria and Archaeans", a chapter of the On-Line Biology Book, provided by Dr M J Farabee of the Estrella Mountain Community College, Arizona, USA at the website www.emc.maricotpa.edu/faculty/farabee/BIOBK/BioBookDiversity. In certain embodiments, the host cell can be a Pseudomonad cell, and can typically be a *P. fluorescens* cell. In other embodiments, the host cell can also be an *E. coli* cell. In another embodiment the host cell can be a eukaryotic cell, for example an insect cell, including but not limited to a cell from a Spodoptera, Trichoplusia, Drosophila or an Estigmene species, or a mammalian cell, including but not limited to a murine cell, a hamster cell, a monkey, a primate or a human cell.

In one embodiment, the host cell can be a member of any of the bacterial taxa. The cell can, for example, be a member of any species of eubacteria. The host can be a member of any one of the taxa: Acidobacteria, Actinobacteira, Aquificae, Bacteroidetes, Chlorobi, Chlamydiae, Choroflexi, Chrysiogenetes, Cyanobacteria, Deferribacteres, Deinococcus, Dictyoglomi, Fibrobacteres, Firmicutes, Fusobacteria, Gemmatimonadetes, Lentisphaerae, Nitrospirae, Planctomycetes, Proteobacteria, Spirochaetes, Thermodesulfobacteria, Thermomicrobia, Thermotogae, Thermus (Thermales), or Verrucomicrobia. In an embodiment of a eubacterial host cell, the cell can be a member of any species of eubacteria, excluding Cyanobacteria.

The bacterial host can also be a member of any species of Proteobacteria. A proteobacterial host cell can be a member of any one of the taxa Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, or Epsilonproteobacteria. In addition, the host can be a member of any one of the taxa Alphaproteobacteria, Betaproteobacteria, or Gammaproteobacteria, and a member of any species of Gammaproteobacteria.

In one embodiment of a Gamma Proteobacterial host, the host will be member of any one of the taxa Aeromonadales, Alteromonadales, Enterobacteriales, Pseudomonadales, or Xanthomonadales; or a member of any species of the Enterobacteriales or Pseudomonadales. In one embodiment, the host cell can be of the order Enterobacteriales, the host cell will be a member of the family Enterobacteriaceae, or may be a member of any one of the genera Erwinia, Escherichia, or Serratia; or a member of the genus Escherichia. Where the host cell is of the order Pseudomonadales, the host cell may be a member of the family Pseudomonadaceae, including the genus Pseudomonas. Gamma Proteobacterial hosts include members of the species Escherichia coli and members of the species *Pseudomonas fluorescens*.

Other Pseudomonas organisms may also be useful. Pseudomonads and closely related species include Gram-negative Proteobacteria Subgroup 1, which include the group of Proteobacteria belonging to the families and/or genera described as "Gram-Negative Aerobic Rods and Cocci" by R. E. Buchanan and N.E. Gibbons (eds.), Bergey's Manual of Determinative Bacteriology, pp. 217-289 (8th ed., 1974) (The Williams & Wilkins Co., Baltimore, Md., USA) (hereinafter "Bergey (1974)"). Table 7 presents these families and genera of organisms.

Table 7. Families and Genera Listed in the Part, "Gram-Negative Aerobic Rods and Cocci" (in Bergey (1974))	
Family I. Pseudomonaceae	Gluconobacter Pseudomonas Xanthomonas Zoogloea
Family II. Azotobacteraceae	Azomonas Azotobacter Beijerinckia Derxia
Family III. Rhizobiaceae	Agrobacterium Rhizobium
Family IV. Methylomonadaceae	Methylococcus Methylomonas
Family V. Halobacteriaceae	Halobacterium Halococcus
Other Genera	Acetobacter Alcaligenes Bordetella Brucella Francisella Thermus

"Gram-negative Proteobacteria Subgroup 1" also includes Proteobacteria that would be classified in this heading according to the criteria used in the classification. The heading also includes groups that were previously classified in this section but are no longer, such as the genera Acidovorax, Brevundimonas, Burkholderia, Hydrogenophaga, Oceanimonas, Ralstonia, and Stenotrophomonas, the genus Sphingomonas (and the genus Blastomonas, derived therefrom), which was created by regrouping organisms belonging to (and previously called species of) the genus Xanthomonas, the genus Acidomonas, which was created by regrouping organisms belonging to the genus Acetobacter as defined in Bergey (1974). In addition hosts can include cells from the genus Pseudomonas, *Pseudomonas enalia* (ATCC 14393), *Pseudomonas nigrifaciensi* (ATCC 19375), and *Pseudomonas putrefaciens* (ATCC 8071), which have been reclassified respectively as *Alteromonas haloplanktis*, *Alteromonas nigrifaciens*, and *Alteromonas putrefaciens*. Similarly, e.g., *Pseudomonas acidovorans* (ATCC 15668) and *Pseudomonas testosteroni* (ATCC 11996) have since been reclassified as *Comamonas acidovorans* and *Comamonas testosteroni*, respectively; and *Pseudomonas nigrifaciens* (ATCC 19375) and *Pseudomonas piscicida* (ATCC 15057) have been reclassified respectively as

Pseudoalteromonas nigrifaciens and *Pseudoalteromonas piscicida*. "Gram-negative Proteobacteria Subgroup 1" also includes Proteobacteria classified as belonging to any of the families: Pseudomonadaceae, Azotobacteraceae (now often called by the synonym, the "Azotobacter group" of Pseudomonadaceae), Rhizobiaceae, and Methylomonadaceae (now often called by the synonym, "Methylococcaceae"). Consequently, in addition to those genera otherwise described herein, further Proteobacterial genera falling within "Gram-negative Proteobacteria Subgroup 1" include: 1) Azotobacter group bacteria of the genus Azorhizophilus; 2) Pseudomonadaceae family bacteria of the genera Cellvibrio, Oligella, and Teredinibacter; 3) Rhizobiaceae family bacteria of the genera Chelatobacter, Ensifer, Liberibacter (also called "Candidatus Liberibacter"), and Sinorhizobium; and 4) Methylococcaceae family bacteria of the genera Methylobacter, Methylocaldum, Methylomicrobium, Methylosarcina, and Methylosphaera.

In another embodiment, the host cell is selected from "Gram-negative Proteobacteria Subgroup 2." "Gram-negative Proteobacteria Subgroup 2" is defined as the group of Proteobacteria of the following genera (with the total numbers of catalog-listed, publicly-available, deposited strains thereof indicated in parenthesis, all deposited at ATCC, except as otherwise indicated): Acidomonas (2); Acetobacter (93); Gluconobacter (37); Brevundimonas (23); Beyerinckia (13); Derxia (2); Brucella (4); Agrobacterium (79); Chelatobacter (2); Ensifer (3); Rhizobium (144); Sinorhizobium (24); Blastomonas (1); Sphingomonas (27); Alcaligenes (88); Bordetella (43); Burkholderia (73); Ralstonia (33); Acidovorax (20); Hydrogenophaga (9); Zoogloea (9); Methylobacter (2); Methylocaldum (1 at NCIMB); Methylococcus (2); Methylomicrobium (2); Methylomonas (9); Methylosarcina (1); Methylosphaera; Azomonas (9); Azorhizophilus (5); Azotobacter (64); Cellvibrio (3); Oligella (5); Pseudomonas (1139); Francisella (4); Xanthomonas (229); Stenotrophomonas (50); and Oceanimonas (4).

Exemplary host cell species of "Gram-negative Proteobacteria Subgroup 2" include, but are not limited to the following bacteria (with the ATCC or other deposit numbers of exemplary strain(s) thereof shown in parenthesis): *Acidomonas methanolica* (ATCC 43581); *Acetobacter aceti* (ATCC 15973); *Gluconobacter oxydans* (ATCC 19357); *Brevundimonas diminuta* (ATCC 11568); *Beijerinckia indica* (ATCC 9039 and ATCC 19361); *Derxia gummosa* (ATCC 15994); *Brucella melitensis* (ATCC 23456), *Brucella abortus* (ATCC 23448); *Agrobacterium*

tumefaciens (ATCC 23308), *Agrobacterium radiobacter* (ATCC 19358), *Agrobacterium rhizogenes* (ATCC 11325); *Chelatobacter heintzii* (ATCC 29600); *Ensifer adhaerens* (ATCC 33212); *Rhizobium leguminosarum* (ATCC 10004); *Sinorhizobium fredii* (ATCC 35423); *Blastomonas natatoria* (ATCC 35951); *Sphingomonas paucimobilis* (ATCC 29837); *Alcaligenes faecalis* (ATCC 8750); *Bordetella pertussis* (ATCC 9797); *Burkholderia cepacia* (ATCC 25416); *Ralstonia pickettii* (ATCC 27511); *Acidovorax facilis* (ATCC 11228); *Hydrogenophagaflava* (ATCC 33667); *Zoogloea ramigera* (ATCC 19544); *Methylobacter luteus* (ATCC 49878); *Methylocaldum gracile* (NCIMB 11912); *Methylococcus capsulatus* (ATCC 19069); *Methylomicrobium agile* (ATCC 35068); *Methylomonas methanica* (ATCC 35067); *Methylosarcina fibrata* (ATCC 700909); *Methylosphaera hansonii* (ACAM 549); *Azomonas agilis* (ATCC 7494); *Azorhizophilus paspali* (ATCC 23833); *Azotobacter chroococcum* (ATCC 9043); *Cellvibrio mixtus* (UQM 2601); *Oligella urethralis* (ATCC 17960); *Pseudomonas aeruginosa* (ATCC 10145), *Pseudomonas fluorescens* (ATCC 35858); *Francisella tularensis* (ATCC 6223); *Stenotrophomonas maltophilia* (ATCC 13637); *Xanthomonas campestris* (ATCC 33913); and *Oceanimonas doudoroffii* (ATCC 27123).

In another embodiment, the host cell is selected from "Gram-negative Proteobacteria Subgroup 3." "Gram-negative Proteobacteria Subgroup 3" is defined as the group of Proteobacteria of the following genera: Brevundimonas; Agrobacterium; Rhizobium; Sinorhizobium; Blastomonas; Sphingomonas; Alcaligenes; Burkholderia; Ralstonia; Acidovorax; Hydrogenophaga; Methylobacter; Methylocaldum; Methylococcus; Methylomicrobium; Methylomonas; Methylosarcina; Methylosphaera; Azomonas; Azorhizophilus; Azotobacter; Cellvibrio; Oligella; Pseudomonas; Teredinibacter; Francisella; Stenotrophomonas; Xanthomonas; and Oceanimonas.

In another embodiment, the host cell is selected from "Gram-negative Proteobacteria Subgroup 4." "Gram-negative Proteobacteria Subgroup 4" is defined as the group of Proteobacteria of the following genera: Brevundimonas; Blastomonas; Sphingomonas; Burkholderia; Ralstonia; Acidovorax; Hydrogenophaga; Methylobacter; Methylocaldum; Methylococcus; Methylomicrobium; Methylomonas; Methylosarcina; Methylosphaera; Azomonas; Azorhizophilus; Azotobacter; Cellvibrio; Oligella; Pseudomonas; Teredinibacter; Francisella; Stenotrophomonas; Xanthomonas; and Oceanimonas.

In another embodiment, the host cell is selected from "Gram-negative Proteobacteria Subgroup 5." "Gram-negative Proteobacteria Subgroup 5" is defined as the group of Proteobacteria of the following genera: Methylobacter; Methylocaldum; Methylococcus; Methylomicrobium; Methylomonas; Methylosarcina; Methylosphaera; Azomonas; Azorhizophilus; Azotobacter; Cellvibrio; Oligella; Pseudomonas; Teredinibacter; Francisella; Stenotrophomonas; Xanthomonas; and Oceanimonas.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 6." "Gram-negative Proteobacteria Subgroup 6" is defined as the group of Proteobacteria of the following genera: Brevundimonas; Blastomonas; Sphingomonas; Burkholderia; Ralstonia; Acidovorax; Hydrogenophaga; Azomonas; Azorhizophilus; Azotobacter; Cellvibrio; Oligella; Pseudomonas; Teredinibacter; Stenotrophomonas; Xanthomonas; and Oceanimonas.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 7." "Gram-negative Proteobacteria Subgroup 7" is defined as the group of Proteobacteria of the following genera: Azomonas; Azorhizophilus; Azotobacter; Cellvibrio; Oligella; Pseudomonas; Teredinibacter; Stenotrophomonas; Xanthomonas; and Oceanimonas.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 8." "Gram-negative Proteobacteria Subgroup 8" is defined as the group of Proteobacteria of the following genera: Brevundimonas; Blastomonas; Sphingomonas; Burkholderia; Ralstonia; Acidovorax; Hydrogenophaga; Pseudomonas; Stenotrophomonas; Xanthomonas; and Oceanimonas.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 9." "Gram-negative Proteobacteria Subgroup 9" is defined as the group of Proteobacteria of the following genera: Brevundimonas; Burkholderia; Ralstonia; Acidovorax; Hydrogenophaga; Pseudomonas; Stenotrophomonas; and Oceanimonas.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 10." "Gram-negative Proteobacteria Subgroup 10" is defined as the group of Proteobacteria of the following genera: Burkholderia; Ralstonia; Pseudomonas; Stenotrophomonas; and Xanthomonas.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 11." "Gram-negative Proteobacteria Subgroup 11" is defined as the group of Proteobacteria of the genera: Pseudomonas; Stenotrophomonas; and Xanthomonas.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 12." "Gram-negative Proteobacteria Subgroup 12" is defined as the group of Proteobacteria of the following genera: Burkholderia; Ralstonia; Pseudomonas. The host cell can be selected from "Gram-negative Proteobacteria Subgroup 13." "Gram-negative Proteobacteria Subgroup 13" is defined as the group of Proteobacteria of the following genera: Burkholderia; Ralstonia; Pseudomonas; and Xanthomonas. The host cell can be selected from "Gram-negative Proteobacteria Subgroup 14." "Gram-negative Proteobacteria Subgroup 14" is defined as the group of Proteobacteria of the following genera: Pseudomonas and Xanthomonas. The host cell can be selected from "Gram-negative Proteobacteria Subgroup 15." "Gram-negative Proteobacteria Subgroup 15" is defined as the group of Proteobacteria of the genus Pseudomonas.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 16." "Gram-negative Proteobacteria Subgroup 16" is defined as the group of Proteobacteria of the following Pseudomonas species (with the ATCC or other deposit numbers of exemplary strain(s) shown in parenthesis): *Pseudomonas abietaniphila* (ATCC 700689); *Pseudomonas aeruginosa* (ATCC 10145); *Pseudomonas alcaligenes* (ATCC 14909); *Pseudomonas anguilliseptica* (ATCC 33660); *Pseudomonas citronellolis* (ATCC 13674); *Pseudomonas flavescens* (ATCC 51555); *Pseudomonas mendocina* (ATCC 25411); *Pseudomonas nitroreducens* (ATCC 33634); *Pseudomonas oleovorans* (ATCC 8062); *Pseudomonas pseudoalcaligenes* (ATCC 17440); *Pseudomonas re*sinovorans (ATCC 14235); *Pseudomonas straminea* (ATCC 33636); *Pseudomonas agarici* (ATCC 25941); *Pseudomonas alcaliphila*; *Pseudomonas alginovora*; *Pseudomonas andersonii*; *Pseudomonas asplenii* (ATCC 23835); *Pseudomonas azelaica* (ATCC 27162); *Pseudomonas beyerinckii* (ATCC 19372); *Pseudomonas borealis*; *Pseudomonas boreopolis* (ATCC 33662); *Pseudomonas brassicacearum*; *Pseudomonas butanovora* (ATCC 43655); *Pseudomonas cellulosa* (ATCC 55703); *Pseudomonas aurantiaca* (ATCC 33663); *Pseudomonas chlororaphis* (ATCC 9446, ATCC 13985, ATCC 17418, ATCC 17461); *Pseudomonas fragi* (ATCC 4973); *Pseudomonas lundensis* (ATCC 49968); *Pseudomonas taetrolens* (ATCC 4683); *Pseudomonas cissicola* (ATCC 33616); *Pseudomonas coronafaciens*; *Pseudomonas diterpeniphila*; *Pseudomonas elongata* (ATCC 10144); *Pseudomonasflectens* (ATCC 12775); *Pseudomonas azotoformans*; *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata* (ATCC 29736); *Pseudomonas extremorientalis*;

Pseudomonas fluorescens (ATCC 35858); *Pseudomonas gessardii*; *Pseudomonas libanensis*; *Pseudomonas mandelii* (ATCC 700871); *Pseudomonas marginalis* (ATCC 10844); *Pseudomonas migulae*; *Pseudomonas mucidolens* (ATCC 4685); *Pseudomonas orientalis*; *Pseudomonas rhodesiae*; *Pseudomonas synxantha* (ATCC 9890); *Pseudomonas tolaasii* (ATCC 33618); *Pseudomonas veronii* (ATCC 700474); *Pseudomonas frederiksbergensis*; *Pseudomonas geniculata* (ATCC 19374); *Pseudomonas gingeri*; *Pseudomonas graminis*; *Pseudomonas grimontii*; *Pseudomonas halodenitrificans*; *Pseudomonas halophila*; *Pseudomonas hibiscicola* (ATCC 19867); *Pseudomonas huttiensis* (ATCC 14670); *Pseudomonas hydrogenovora*; *Pseudomonas jessenii* (ATCC 700870); *Pseudomonas kilonensis*; *Pseudomonas lanceolata* (ATCC 14669); *Pseudomonas lini*; *Pseudomonas marginata* (ATCC 25417); *Pseudomonas mephitica* (ATCC 33665); *Pseudomonas denitrificans* (ATCC 19244); *Pseudomonas pertucinogena* (ATCC 190); *Pseudomonas pictorum* (ATCC 23328); *Pseudomonas psychrophila*; *Pseudomonas filva* (ATCC 31418); *Pseudomonas monteilii* (ATCC 700476); *Pseudomonas mosselii*; *Pseudomonas oryzihabitans* (ATCC 43272); *Pseudomonas plecoglossicida* (ATCC 700383); *Pseudomonas putida* (ATCC 12633); *Pseudomonas reactans*; *Pseudomonas spinosa* (ATCC 14606); *Pseudomonas balearica*; *Pseudomonas luteola* (ATCC 43273);. *Pseudomonas stutzeri* (ATCC 17588); *Pseudomonas amygdali* (ATCC 33614); *Pseudomonas avellanae* (ATCC 700331); *Pseudomonas caricapapayae* (ATCC 33615); *Pseudomonas cichorii* (ATCC 10857); *Pseudomonas ficuserectae* (ATCC 35104); *Pseudomonas fuscovaginae*; *Pseudomonas meliae* (ATCC 33050); *Pseudomonas syringae* (ATCC 19310); *Pseudomonas viridiflava* (ATCC 13223); *Pseudomonas thermocarboxydovorans* (ATCC 35961); *Pseudomonas thermotolerans*; *Pseudomonas thivervalensis*; *Pseudomonas vancouverensis* (ATCC 700688); *Pseudomonas wisconsinensis*; and *Pseudomonas xiamenensis*.

The host cell can be selected from "Gram-negative Proteobacteria Subgroup 17." "Gram-negative Proteobacteria Subgroup 17" is defined as the group of Proteobacteria known in the art as the "fluorescent Pseudomonads" including those belonging, e.g., to the following Pseudomonas species: *Pseudomonas azotoformans*; *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata*; *Pseudomonas extremorientalis*; *Pseudomonas fluorescens*; *Pseudomonas gessardii*; *Pseudomonas libanensis*; *Pseudomonas mandelii*; *Pseudomonas marginalis*;

Pseudomonas migulae; *Pseudomonas mucidolens*; *Pseudomonas orientalis*; *Pseudomonas rhodesiae*; *Pseudomonas synxantha*; *Pseudomonas tolaasii*; and *Pseudomonas veronii*.

Other suitable hosts include those classified in other parts of the reference, such as Gram (+) Proteobacteria. In one embodiment, the host cell is an *E. coli*. The genome sequence for *E. coli* has been established for *E. coli* MG1655 (Blattner, *et al.* (1997) The complete genome sequence of Escherichia coli K-12, *Science* 277(5331): 1453-74) and DNA microarrays are available commercially for *E. coli* K12 (MWG Inc, High Point, N.C.). *E. coli* can be cultured in either a rich medium such as Luria-Bertani (LB) (10 g/L tryptone, 5 g/L NaCl, 5 g/L yeast extract) or a defined minimal medium such as M9 (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4C1, 0.5 g/L NaCl, pH 7.4) with an appropriate carbon source such as 1% glucose. Routinely, an over night culture of *E. coli* cells is diluted and inoculated into fresh rich or minimal medium in either a shake flask or a fermentor and grown at 37°C.

A host cell can also be of mammalian origin, such as a cell derived from a mammal including any human or non-human mammal. Mammals can include, but are not limited to primates, monkeys, porcine, ovine, bovine, rodents, ungulates, pigs, swine, sheep, lambs, goats, cattle, deer, mules, horses, monkeys, apes, dogs, cats, rats, and mice.

A host cell may also be of plant origin. Cells from any plant can be selected in which to screen for the production of a heterologous protein of interest. Examples of suitable plant include, but are not limited to, alfalfa, apple, apricot, Arabidopsis, artichoke, arugula, asparagus, avocado, banana, barley, beans, beet, blackberry, blueberry, broccoli, brussels sprouts, cabbage, canola, cantaloupe, carrot, cassaya, castorbean, cauliflower, celery, cherry, chicory, cilantro, citrus, clementines, clover, coconut, coffee, corn, cotton, cranberry, cucumber, Douglas fir, eggplant, endive, escarole, eucalyptus, fennel, figs, garlic, gourd, grape, grapefruit, honey dew, jicama, kiwifruit, lettuce, leeks, lemon, lime, Loblolly pine, linseed, mango, melon, mushroom, nectarine, nut, oat, oil palm, oil seed rape, okra, olive, onion, orange, an ornamental plant, palm, papaya, parsley, parsnip, pea, peach, peanut, pear, pepper, persimmon, pine, pineapple, plantain, plum, pomegranate, poplar, potato, pumpkin, quince, radiata pine, radiscchio, radish, rapeseed, raspberry, rice, rye, sorghum, Southern pine, soybean, spinach, squash, strawberry, sugarbeet, sugarcane, sunflower,

sweet potato, sweetgum, tangerine, tea, tobacco, tomato, triticale, turf, turnip, a vine, watermelon, wheat, yams, and zucchini. In some embodiments, plants useful in the method are Arabidopsis, corn, wheat, soybean, and cotton.

Kits

The present invention also provides kits useful for identifying a host strain, e.g. a *P. fluorescens* host strain, optimal for producing a heterologous protein or polypeptide of interest. The kit comprises a plurality of phenotypically distinct host cells, wherein each population has been genetically modified to increase the expression of one or more target genes involved in protein production, to decrease the expression of one or more target genes involved in protein degradation, or both. The array may further comprise one or more populations of cells that have not been genetically modified to modulate the expression of either a gene involved in protein production or a gene involved in protein degradation. These kits may also comprise reagents sufficient to facilitate growth and maintenance of the cell populations as well as reagents and/or constructs for expression of a heterologous protein or polypeptide of interest. The populations of host cells may be provided in the kit in any manner suitable for storage, transport, and reconstitution of cell populations. The cell populations may be provided live in a tube, on a plate, or on a slant, or may be preserved either freeze-dried or frozen in a tube or vial. The cell populations may contain additional components in the storage media such as glycerol, sucrose, albumin, or other suitable protective or storage agents.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL EXAMPLES

Overview

Heterologous protein production often leads to the formation of insoluble or improperly folded proteins, which are difficult to recover and may be inactive. Furthermore, the presence of specific host cell proteases may degrade the protein of interest and thus reduce the final yield. There is no single factor that will improve the production of all heterologous proteins. Thus, a method was sought to identify factors specific to a particular heterologous protein from a pool of likely candidates.

Using Systems Biology tools, the *P. fluorescens* genome was mined to identify host cell protein folding modulator and protease genes. Then, global gene expression analyses were performed to prioritize upregulated targets, and, thereafter, novel protein production strains were constructed. As a result, a "Pfēnex Strain Array" was assembled consisting of a plurality of phenotypically distinct P. fluorescens host strains that are deficient in host-cell proteases or allow the co-overexpression of protein folding modulators. This strain array can be used to screen for factors that specifically enhance the yield or quality of certain heterologous proteins. Providing a plurality of phenotypically distinct host strains increases the chance of success of identifying a host strain that will increase the production of any individual heterologous protein of interest.

This invention provides an improvement in the production of heterologous proteins in *Pseudomonas fluorescens*. Having available a library of host strains in the same genetic background allows the rapid screening and identification of factors that increase the yield and/or quality of heterologously expressed proteins. The genome sequence of P. fluorescens has been annotated and targeted host cell folding modulators and proteases have been identified. Folding modulators assist in the proper folding of proteins and include chaperones, chaperonins, peptidyl-proline isomerases (PPIases), and disulfide bond formation proteins. Proteases can degrade the protein of interest and thus affect heterologous protein yield and quality. Using background knowledge from the literature and DNA microarray analyses to identify likely targets, a list of about 80 target genes was assembled. In host cells that have the same genetic background, these genes were either removed from the genome or cloned into plasmids to enable co-overexpression along with heterologous proteins. The resulting strains were arrayed in 96-well format and, after transformation of plasmids that express the heterologous protein of interest, were screened for improved protein yield and/or quality.

Example 1

Identification of Folding Modulator Genes in the Genome of P. fluorescens Strain MB214

Folding modulators are a class of proteins present in all cells which aid in the folding, unfolding and degradation of nascent and heterologous polypeptides. Folding modulators include chaperones, chaperonins, peptidyl-prolyl *cis-trans* isomerases, and

proteins involved in protein disulfide bond formation. As a first step to construct novel production strains with the ability to help fold heterologous proteins, the *P. fluorescens* genome was mined to identify host cell folding modulator genes.

Each of the 6,433 predicted ORFs of the *P. fluorescens* MB214 genome was analyzed for the possibility that they encoded a folding modulator using the following method. Several folding modulators of interest had already been identified by Dow researchers by analysis of the genome annotation (Ramseier et. al. 2001). Homologs of these starting proteins were identified using protein/protein BLAST with the starting protein as the query and a database of all MB214 translated ORFs as the subject. Those translated ORFs which matched the query proteins with significant homology were added to the list for further analysis. Significant homology is defined here as having an e-score of 1e-30 or less with allowances made for human judgment based on the length and quality of the alignment. The intention of this study was to be very inclusive to maximize the chance that all potential folding modulators would be identified.

More ORFs were added to the list based on their curated function from the previous annotation containing the keyword "chaperone". Finally, the ORFs were analyzed by the protein signature family searching program InterProScan (Quevillon et. al. 2005) against the InterPro Database version 7.0 (Mulder et. al. 2005). The ORFs were assigned protein families by the InterProScan software as well as Gene Ontology (GO) categories associated with those families (Gene Ontology Consortium. 2004). Using these automatic GO assignments, all of the ORFs which had been assigned the GO terms "GO:0006457 Biological Process: protein folding" or "GO:0003754 Molecular Function: chaperone activity" were added to the list for further analysis.

The list was then analyzed to remove ORFs which had a low probability of encoding folding modulators. Again, the intent of this study was to be very inclusive but many of the ORFs assigned to the list by these semi-automated methods could be easily identified as not coding for folding modulators based on limited criteria and human judgment.

The most common reason for excluding a certain ORF was the weak evidence that this ORF is actually a folding modulator, i.e. ORFs which had been assigned to the list based on the previous annotation where the reasoning for annotating the ORF as a folding modulator was either unclear or contradictory. InterProScan is actually a

conglomerate of different programs and some of these programs are considered to be more reliable than others. If an ORF was assigned to the list based solely on the output of the ScanRegExp or ProfileScan components then it was removed. The final list of *P. fluorescens* folding modulators has 43 members and is shown in Table 1.

Example 2

Identification of Protease Genes in the Genome of P. fluorescens Strain MB214

Proteases are enzymes that hydrolyze peptide bonds and are necessary for the survival of all living creatures. However, their role in the cell means that proteases can be detrimental to recombinant protein yield and/or quality in any heterologous protein expression system, which also includes the Pfenex Expression Technology™. As a first step to construct novel production strains that have protease genes removed from the genome, the *P. fluorescens* genome was mined to identify host cell protease genes.

Each of the 6,433 predicted ORFs of the *P. fluorescens* MB214 genome were analyzed for the possibility that they encoded a protease using the following method. The MEROPS database is manually curated by researchers at the Wellcome Trust Sanger Institute, Cambridge, UK (Rawlings *et. al.* 2006, http://merops.sanger.ac.uk). It is a comprehensive list of proteases discovered both through laboratory experiments as well as by homology to known protease families. One of the strengths of the database is the MEROPS hierarchical classification scheme. In this system, homologs which share the same function are grouped together into families. Families are grouped into clans based on evolutionary relatedness that again are based on similar structural characteristics. The method makes great use of the database to identify protease homologs within the *P. fluorescens* genome.

Homologs to the MEROPS database were identified using protein/protein BLAST with each MB214 translated ORF as the query and a database of all of the MEROPS proteins as the subject. Those translated ORFs, which matched the query proteins with significant homology, were added to the list for further analysis. Significant homology in this case is defined here as having an e-score of $1e^{-60}$ or less with allowances made for human judgment based on the length and quality of the alignment. This step yielded 109 potential proteases for the list.

The ORFs were also analyzed by the protein signature family searching program InterProScan (Quevillon *et. al.* 2005) against the InterPro Database version 7.0 (Mulder *et. al.* 2005). The ORFs were assigned protein families by the InterProScan software as well as Gene Ontology (GO) categories associated with those families (Gene Ontology Consortium. 2004). Using these automatic GO assignments, all of the ORFs which had been assigned a GO name that contained the strings "peptidase", "protease" or "proteolysis" were added to the list for further analysis. This step yielded an additional 70 potential proteases that had not been identified in the previous step.

More ORFs were added to the list based on their curated function from the previous annotation (Ramseier *et. al.* 2001) containing the keywords "peptidase" or "protease". This step yielded 32 potential proteases that again had not been identified in the previous steps.

The list was then analyzed to remove ORFs which had a low probability of encoding proteases. Again, the intent of this study was to be very inclusive but many of the ORFs assigned to the list by these semi-automated methods could be easily identified as not coding for proteases based on limited criteria and human judgment. The two most common reasons for excluding genes were the weak evidence that a certain ORF is actually a protease, or that a particular gene showed greatest homology with another protein known to be protease homolog but not a protease itself. The final list of *P. fluorescens* proteases has 90 members and is shown in Table 2.

Example 3

In Silico Cellular Location Prediction of the Folding Modulator and Protease Proteins

One of the strengths of the Pfenex Expression Technology™ is its ability to control the cellular compartment to which a particular heterologous protein can be segregated. Thus, the cellular compartments where the identified host cell folding modulator and protease proteins are located were predicted. To make these predictions, two programs were chosen. PsortB 2.0 combines the results of 12 separate algorithms, which predict the subcellular location of a given peptide. The majority of the algorithms rely on detecting homology between the query protein and proteins of known subcellular localization. PsortB also includes algorithms such as

HMMTOP and SignalP, which detect the presence of transmembrane folding domains or type I secretion signal sequences, respectively, using Hidden Markov Models (HMM). In addition to the PsortB results, SignalP HMM was used to predict the presence of type I secretion signal sequences. This was necessary because the output of PsortB can be vague when a signal sequence is detected but no other specific information indicating the subcellular location is given. In these cases, PsortB indicates that the subcellular localization of the protein is unknown, because it really could segregate to any one of the cytoplasmic membrane, periplasm, outer membrane or extracellular compartments. However, it is informative enough to know that the protein is probably not located in the cytoplasm to make it worth noting that in the table. Thus, Table 2 lists the results of the PsortB algorithm except in cases where that result was unknown. In these cases the result of SignalP HMM alone is given with "Signal Peptide" indicating that a signal peptide was detected and "Non Secretory" indicating that no signal peptide was detected.

Example 4

Construction of Plasmids that enable the Co-Overexpression of Folding Modulators

Folding modulator genes were cloned into a plasmid derivative of pCN (Nieto et al. 1990), which is compatible with another plasmid that routinely is used to express the heterologous protein of interest (Squires et al. 2004; Chew et al. 2005). The construction of a mannitol-inducible *grpE-dnaKJ*-containing plasmid is exemplified. Other folding modulators – either as a single gene or as multiple genes when organized in operons – were cloned similarly as outlined below.
Employing genomic DNA isolated from *P. fluorescens* MB214 (DNeasy; Qiagen, Valencia, CA) as a template and primers RC199 (5-ATATACTAGTAGGAGGTAACTTATGGCTGACGAACAGACGCA-3') (SEQ ID NO:1) and RC200 (5'-ATATTCTAGATTACAGGTCGCCGAAGAAGC-3') (SEQ ID NO:2), the *grpE-dnaKJ* genes were amplified using *PfuTurbo* (Stratagene, La Jolla, CA) as per the manufacturer's recommendations. The resulting 4 kb PCR product was digested with *Spe*I and *Xba*I (restriction sites underlined in the primers above) and ligated into pDOW2236 which is a derivative of pDOW1306-6 (Schneider et al. 2005b) to create pDOW2240 containing the *grpE-dnaKJ* operon under control of the *tac* promoter. Plasmid pDOW2240 was then digested with *Spe*I and *Hind*III and the resulting *grpE-*

dnaKJ-containing 4.0 kb DNA fragment was gel-purified using Qiaquick (Qiagen, Valencia, CA) and ligated into pDOW2247, which is a derivative of pCN carrying the *P. fluorescens* mannitol-regulated promoter (Schneider et al. 2005a), that was also digested with *Spe*I and *Hind*III. The resulting plasmid, pDOW3501, contained the *grpE-dnaKJ* operon under the control of the mannitol promoter. Plasmid pDOW3501 was then transformed into DC388 and other uracil-auxotrophic strains by selecting on M9 glucose plates supplemented with 250 ug/ml uracil.

Example 5

Construction of P. fluorescens Strains with Genomic Deletions of Protease Genes

Plasmids that enabled the creation of genomic deletions were constructed by amplification of 500-1000 bp DNA fragments both 5' and 3' of the gene to be deleted. The resulting 5' PCR product typically ends with the translational initiation codon (ATG or GTG or TGT) of the gene to be deleted while the 3' PCR product typically begins with the stop codon (TAA or TGA or TAG) of the gene to be deleted. These two PCR products were fused together through an additional amplification step then cloned into pDOW1261 (Figure 1) (Chew et al. 2005) using SOE PCR (Horton et al. 1990).

Example 6

High-through-put Growth and Analysis of Heterologous Protein Expression in P. fluorescens Strains

Plasmid pDOW2787 encodes the monoclonal antibody (mAb) gal2; the heavy chain is expressed with a Pbp secretion leader and under control of the *tac* promoter. The light chain is expressed with an OprF secretion leader and under control of the mannitol promoter. The plasmid was electroporated into competent cells of 63 strains carrying either a directed gene deletion or pDOW2247 carrying a folding modulator for co-expression, and five control strains containing a wild type strain. Cells were cultured in replicate deep-well blocks containing growth medium with glycerol by shaking at 300 rpm. Protein expression was induced at 24 hrs with 0.1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and 1% mannitol. At 24 hrs post-induction, aliquots were lysed, antigen-binding of the antigen was measured to quantitate amounts of active antibody. The value was divided by OD_{600} to measure cell specific activity. Strains Δ*prc1, ΔdegP2*, *ΔLa2*, *ΔclpP*, and *Δprc2*, Δ*prc2,* the *grpEdnaKJ* co-

expression strain, Δ*tig, ΔclpX, and* Δ*lon* were all 2.4-fold or more higher than the control strains, which was statistically significant ($p<0.5$). Soluble cells fractions were prepared from Δ*prc1, ΔdegP2*, *ΔLa2* and the *grpEdnaKJ* co-expression strain and subjected to Western analysis (Figure 2). A band with a size consistent with fully assembled antibody was detected in the four test strains, but not in the control.

References

Chew, L. C., T. M. Ramseier, D. M. Retallack, J. C. Schneider, C. H. Squires and H. W. Talbot (2005). Pseudomonas fluorescens. Production of Recombinant Proteins. Novel Microbial and Eucaryotic Expression Systems. G. Gellissen. Weinheim, WILEY-VCH: 45-66

Dolinski, K, Heitman, J. 1997. Peptidyl-prolyl isomerases – an overview of the cyclophilin, FKBP and parvulin families. in Guidebook to Molecular Chaperones and Protein-Folding Catalysts. Gething M-J Ed. Oxford University Press Inc., New York:359-369

Gardy, J.L., M.R. Laird, F. Chen, S. Rey, C.J. Walsh, M. Ester, and F.S.L. Brinkman 2005 PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 21(5):617-623.

Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research 32:D258-D261.

Gething M-J Ed. 1997. Guidebook to Molecular Chaperones and Protein-Folding Catalysts. Oxford University Press Inc., New York.

Horton, R. M., Z. Cai, S. N. Ho and L. R. Pease (1990). "Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction." BioTechniques **8**(5): 528-30, 532, 534-5

Lombardo, M-J, Thanassi, DG, Hultgren, SJ. 1997. Escherichia coli PapD. in Guidebook to Molecular Chaperones and Protein-Folding Catalysts. Gething M-J Ed. Oxford University Press Inc., New York:463-465

Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L, Copley R, Courcelle E, Das U, Durbin R, Fleischmann W, Gough J, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McDowall J, Mitchell A, Nikolskaya AN, Orchard S, Pagni M, Ponting CP, Quevillon E, Selengut

J, Sigrist CJ, Silventoinen V, Studholme DJ, Vaughan R, Wu CH. 2005. InterPro, Progress and Status in 2005. Nucleic Acids Res. 33, Database Issue:D201-5.

Nieto, C., E. Fernandez-Tresguerres, N. Sanchez, M. Vicente and R. Diaz (1990). "Cloning vectors, derived from a naturally occurring plasmid of Pseudomonas savastanoi, specifically tailored for genetic manipulations in Pseudomonas." Gene 87(1): 145-9.

Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. (2005) InterProScan: protein domains identifier. Nucleic Acids Research 33: W116-W120.

Ramseier TM, S. C., Payne J, Chew L, Rothman LD, Subramanian M. 2001. The Pseudomonas fluorescens MB214 Genome Sequence. CRI CRI2001001442; BIOTECH 01-007. The Dow Chemical Company.

Ranson, NA, White, HE, Saibil, HR. 1998. Chaperonins Biochem. J. 333, 233-242.

Rawlings, N.D., Morton, F.R. & Barrett, A.J. 2006. *MEROPS*: the peptidase database. Nucleic Acids Res 34, D270-D272.

Schneider, J. C., A. F. Jenings, D. M. Mun, P. M. McGovern and L. C. Chew (2005a). "Auxotrophic markers pyrF and proC can replace antibiotic markers on protein production plasmids in high-cell-density Pseudomonas fluorescens fermentation." Biotechnology Progress 21(2): 343-348.

Schneider, J. C., B. Rosner and A. Rubio (2005b). Mannitol Induced Promoter Systems in Bacterial Host Cells. USA, The Dow Chemical Company.

Squires, C. H., D. M. Retallack, L. C. Chew, T. M. Ramseier, J. C. Schneider and H. W. Talbot (2004). "Heterologous protein production in P. fluorescens." BioProcess International 2(11): 54-56, 58-59

Table of SEQ ID NO:

PROTEIN FOLDING MODULATOR (RXF#)	SEQ ID NO:	PROTEASE (RXF#)	SEQ ID NO:
RXF02095.1	3	RXF00133.1	46
RXF06767.1	4	RXF00285.2	47
RXF01748.1	5	RXF00325.1	48
RXF03385.1	6	RXF00428.1	49
RXF05399.1	7	RXF00449.1	50
RXF06954.1	8	RXF00458.2	51
RXF03376.1	9	RXF00561.2	52
RXF03987.2	10	RXF00670.1	53
RXF05406.2	11	RXF00811.1	54
RXF03346.2	12	RXF01037.1	55
RXF05413.1	13	RXF01181.1	56
RXF04587.1	14	RXF01250.2	57
RXF08347.1	15	RXF01291.2	58
RXF04654.2	16	RXF01418.1	59
RXF04663.1	17	RXF01590.2	60
RXF01957.2	18	RXF01816.1	61
RXF01961.2	19	RXF01822.2	62
RXF04254.2	20	RXF01918.1	63
RXF05455.2	21	RXF01919.1	64
RXF02231.1	22	RXF01961.2	65
RXF07017.2	23	RXF01968.1	66
RXF08657.2	24	RXF02003.2	67
RXF01002.1	25	RXF02151.2	68
RXF03307.1	26	RXF02161.1	69
RXF04890.2	27	RXF02342.1	70
RXF03768.1	28	RXF02492.1	71
RXF05345.2	29	RXF02689.2	72
RXF06034.2	30	RXF02739.1	73
RXF06591.1	31	RXF02796.1	74
RXF05753.2	32	RXF02980.1	75
RXF01833.2	33	RXF03065.2	76
RXF04655.2	34	RXF03329.2	77
RXF05385.1	35	RXF03364.1	78
RXF00271.1	36	RXF03397.1	79
RXF06068.1	37	RXF03441.1	80
RXF05719.1	38	RXF03488.2	81
RXF03406.2	39	RXF03699.2	82
RXF04296.1	40	RXF03916.1	83
RXF04553.1	41	RXF04047.2	84
RXF04554.2	42	RXF04052.2	85
RXF05310.2	43	RXF04304.1	86
RXF05304.1	44	RXF04424.2	87

RXF05073.1	45	RXF04495.2	88
RXF02090	137	RXF04500.1	89
		RXF04567.1	90
		RXF04631.2	91
		RXF04653.2	92
		RXF04657.2	93
		RXF04663.1	94
		RXF04692.1	95
		RXF04693.1	96
		RXF04715.1	97
		RXF04802.1	98
		RXF04808.2	99
		RXF04920.1	100
		RXF04923.1	101
		RXF04960.2	102
		RXF04968.2	103
		RXF04971.2	104
		RXF05081.1	105
		RXF05113.2	106
		RXF05137.1	107
		RXF05236.1	108
		RXF05379.1	109
		RXF05383.2	110
		RXF05400.2	111
		RXF05615.1	112
		RXF05817.1	113
		RXF05943.1	114
		RXF06281.1	115
		RXF06308.2	116
		RXF06399.2	117
		RXF06451.1	118
		RXF06564.1	119
		RXF06586.1	120
		RXF06755.2	121
		RXF06993.2	122
		RXF07170.1	123
		RXF07210.1	124
		RXF07879.1	125
		RXF08136.2	126
		RXF08517.1	127
		RXF08627.2	128
		RXF08653.1	129
		RXF08773.1	130
		RXF08978.1	131
		RXF09091.1	132
		RXF09147.2	133

		RXF09487.1	134
		RXF09831.2	135
		RXF04892.1	136

All publications and patent applications mentioned in the specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

THAT WHICH IS CLAIMED:

1. An array comprising at least a first and a second population of *Pseudomonas fluorescens* (*P. fluorescens*) cells, wherein each population is selected from the group consisting of:

a) at least one population of *P. fluorescens* cells that has been genetically modified to reduce the expression of at least one target gene involved in protein degradation;

b) at least one population of *P. fluorescens* cells that has been genetically modified to increase the expression of at least one target gene involved in protein production; and,

c) at least one population of *P. fluorescens* cells that has been genetically modified according to (a) and (b);

wherein each population in the array is non-identical and wherein each population is physically separate one from another.

2. The array of claim 1, wherein said at least one target gene involved in protein degradation is a protease.

3. The array of claim 2, wherein the at least one protease is selected from the proteases set forth in Table 2.

4. The array of claim 1, wherein the at least one target gene involved in protein production is a gene involved in proper protein transcription, translation, processing, or translocation of a heterologous protein.

5. The array of claim 4, wherein the at least one target gene is a protein folding modulator.

6. The array of claim 5, wherein the protein folding modulator is selected from the protein folding modulators set forth in Table 1.

7. The array of any of claims 1 to 6, further comprising at least one population of *P. fluorescens* cells that has not been genetically modified to increase the expression of said target gene involved in protein production or said target gene involved in protein degradation.

8. The array of any of claims 1 to 7, wherein at least one population has been genetically modified to increase the expression of two or more target genes involved in protein production.

9. The array of any of claims 1 to 7, wherein at least one population has been genetically modified to decrease the expression of two or more target genes involved in protein degradation.

10. The array of any of claims 1 to 9, wherein said array is in a format conducive to high throughput screening of said array.

11. The array of claim 10, wherein said format is a 96 well format.

12. The array of any of claims 1 to 11, wherein at least one population is capable of expressing at least one heterologous protein of interest when transformed with an expression construct comprising at least one gene encoding the at least one heterologous protein of interest.

13. A kit comprising the array of any of claims 1 to 12.

14. A method for identifying an optimal *Pseudomonas fluorescens* (*P. fluorescens*) host cell for expression of at least one heterologous protein of interest comprising:

a) obtaining an array comprising at least a first and a second population of *P. fluorescens* cells, wherein each population is selected from the group consisting of:

i) at least one population of *P. fluorescens* cells that has been genetically modified to reduce the expression of at least one target gene involved in protein degradation;

ii) at least one population of *P. fluorescens* cells that has been genetically modified to increase the expression of at least one target gene involved in protein production; and,

iii) at least one population of *P. fluorescens* cells that has been genetically modified according to (i) and (ii);

wherein each population in the array is non-identical and wherein each population is physically separate one from another;

b) introducing into at least one cell of each population an expression cassette comprising at least one gene encoding the at least one heterologous protein of interest;

c) maintaining said cells under conditions sufficient for the expression of said protein of interest in at least one population of cells; and,

d) selecting the optimal population of cells in which the heterologous protein of interest is produced;

wherein the heterologous protein of interest exhibits one or more of improved expression, improved activity, improved solubility, improved translocation, or reduced proteolytic degradation in the optimal population of cells compared to other populations in the array.

15. The method of claim 14, wherein said at least one target gene involved in protein degradation is a protease.

16. The method of claim 15, wherein said protease is selected from the proteases set forth in Table 2.

17. The method of claim 14, wherein the at least one target gene involved in protein production is a gene involved in proper protein transcription, translation, processing, or translocation of a heterologous protein.

18. The method of claim 17, wherein the at least one target gene is a protein folding modulator.

19. The method of claim 18, wherein the protein folding modulator is selected from the protein folding modulators set forth in Table 1.

20. The method of any of claims 14 to 19, wherein the array further comprises at least one population of *P. fluorescens* cells that has not been genetically modified to increase the expression of said target gene involved in protein production or said target gene involved in protein degradation.

21. The method of any of claims 14 to 20, wherein at least one population has been genetically modified to increase the expression of two or more target genes involved in protein production.

22. The method of any of claims 14 to 20, wherein at least one population has been genetically modified to decrease the expression of two or more target genes involved in protein degradation.

23. The method of any of claims 14 to 22, wherein said array is in a format conducive to high throughput screening of said array.

24. The method of claim 23, wherein said format is a 96 well format.

25. The method of claim 14, wherein said expression cassette further comprises a signaling peptide operably linked to the heterologous protein of interest.

26. The method of claim 19, wherein said signaling peptide is a secretion signal peptide.

27. The method of claim 19, wherein said signaling peptide is native to the *Pseudomonas fluorescens* host cell.

28. A method for identifying an optimal *Pseudomonas fluorescens* (*P. fluorescens)* host cell for expression of a heterologous protein of interest comprising:

a) obtaining an array comprising at least a first and a second population of *P. fluorescens* cells, wherein each population has been genetically modified to reduce the expression of at least one protein involved in protein degradation, wherein each population in the array is non-identical, and wherein each population is physically separate one from another;

b) lysing the cells;

c) contacting the lysate from each population of cells with the heterologous protein of interest; and,

d) selecting the optimal population of cells;

wherein the optimal population of cells is the population which results in a decreased level of degradation of the heterologous protein of interest in step (c) compared to other populations in the array.

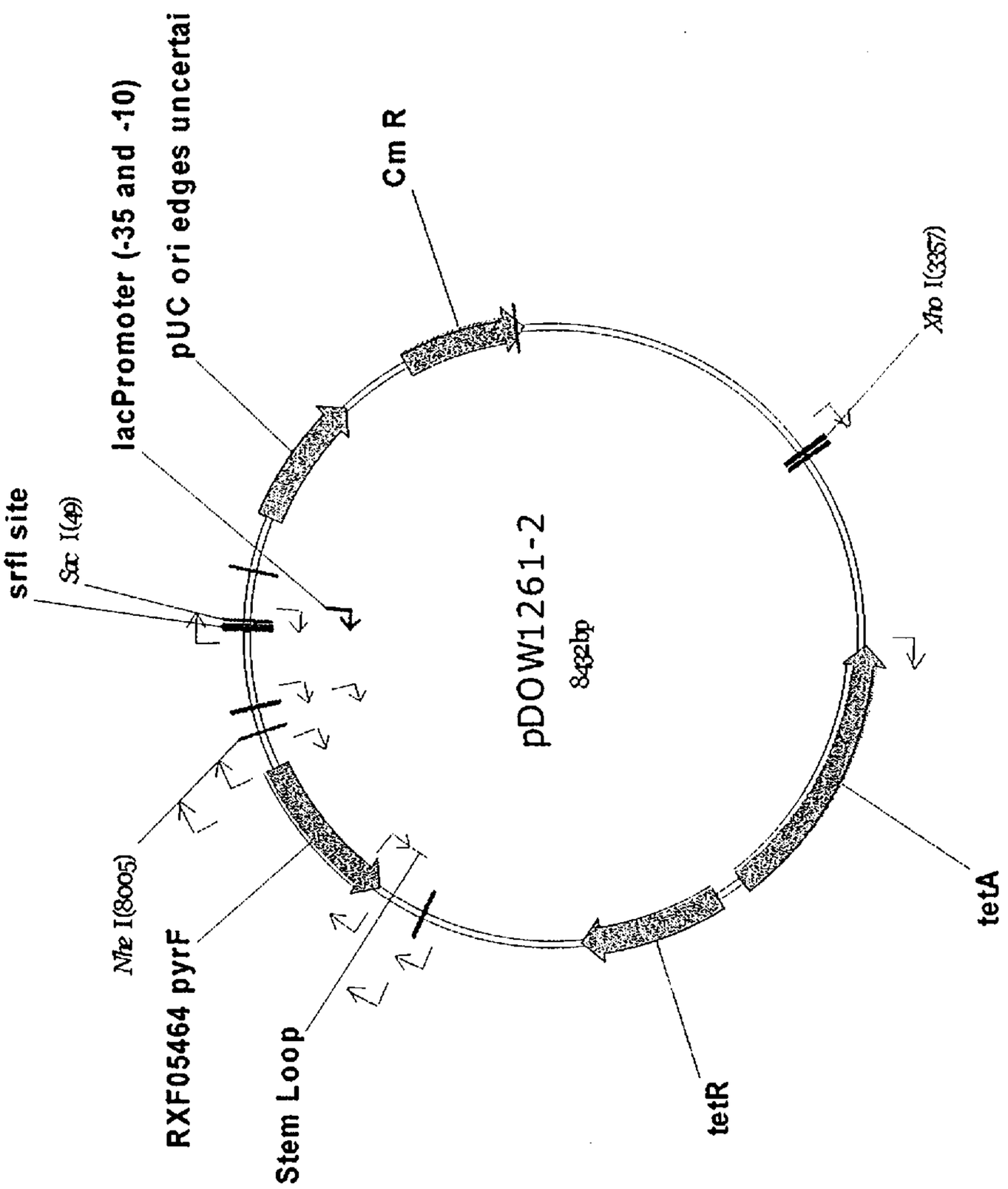
lacPromoter (-35 and -10)
pUC ori edges uncertai
Cm R
srfI site
Sac I(49)
Xho I(3357)
pDOW1261-2
8432bp
Nhe I(8005)
RXF05464 pyrF
Stem Loop
tetR
tetA

FIG. 1A

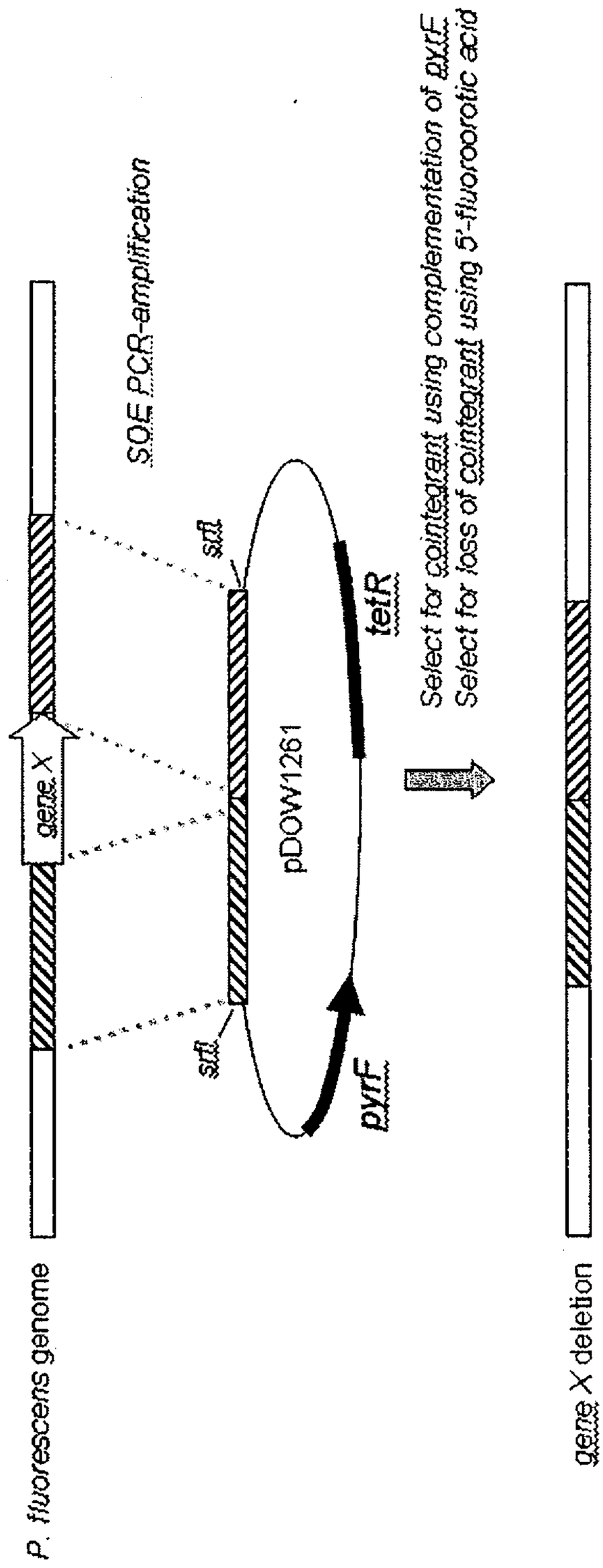
P. fluorescens genome
gene X
SOE PCR-amplification
stfl
pDOW1261
tetR
pyrF
stfl
Select for cointegrant using complementation of pyrF
Select for loss of cointegrant using 5'-fluoroorotic acid
deletion
gene X

FIG. 1B

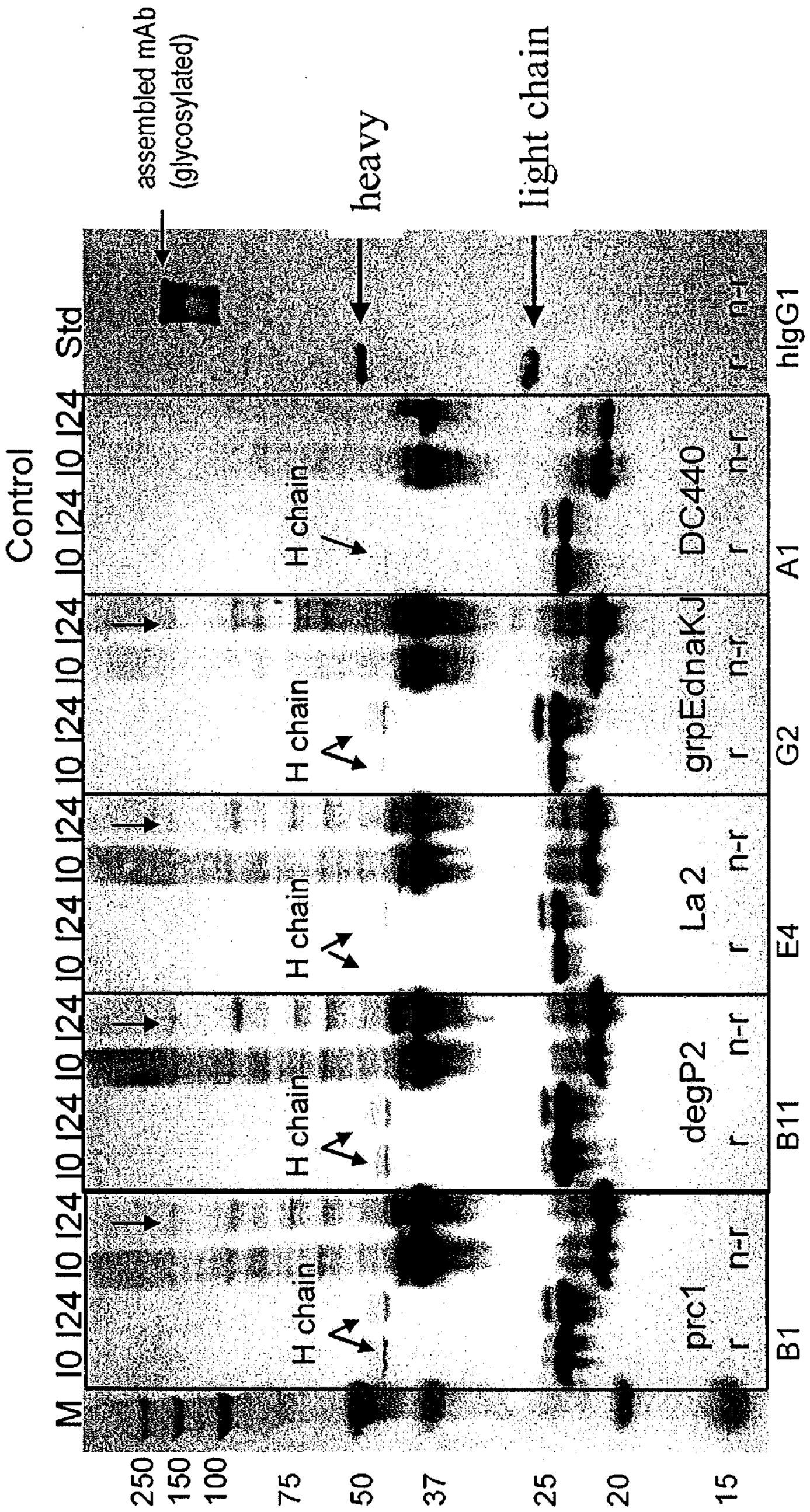
assembled mAb
(glycosylated)
heavy
light chain
Control
Std
I0 I24 I0 I24
H chain
hIgG1
DC440
A1
grpEdnaKJ
G2
La 2
E4
degP2
B11
prc1
B1
M
250
150
100
75
50
37
25
20
15

FIG. 2

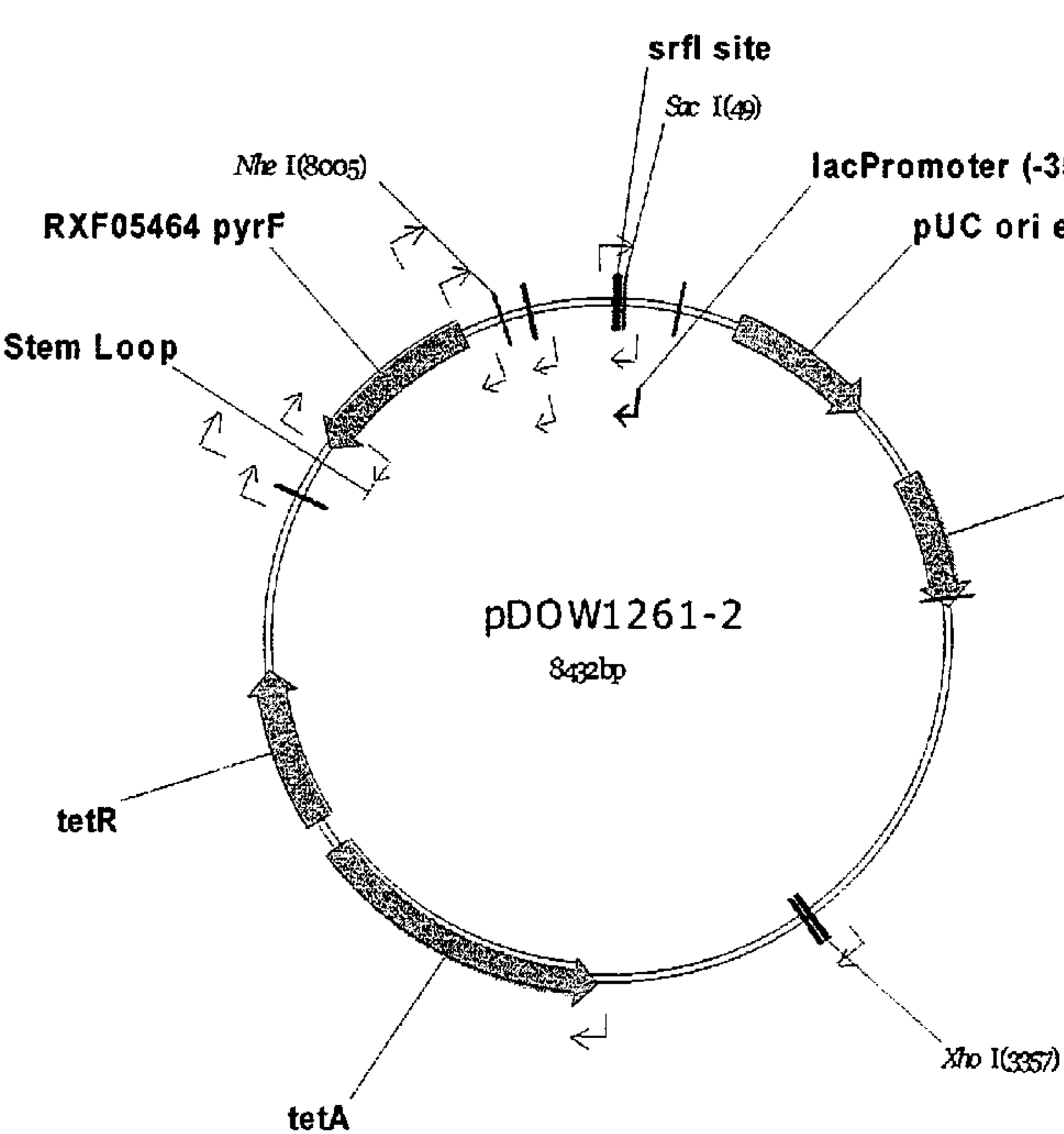
srfI site
Sac I(49)
Nhe I(8005)
lacPromoter (-35 and -10)
RXF05464 pyrF
pUC ori edges uncertai
Stem Loop
Cm R
pDOW1261-2
8432bp
tetR
Xho I(3357)
tetA

FIG. 1A