

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 February 2008 (14.02.2008)

PCT

(10) International Publication Number
WO 2008/019183 A2

(51) International Patent Classification:
C12P 21/06 (2006.01)

(74) Agents: MYCROFT, Frank J. et al.; TOWNSEND and TOWNSEND and CREW LLP, Two Embarcadero Center, Eighth Floor, San Francisco, California 94111 (US).

(21) International Application Number:
PCT/US2007/069300

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, 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.

(22) International Filing Date: 18 May 2007 (18.05.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/801,758 18 May 2006 (18.05.2006) US

(71) Applicant (for all designated States except US): THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 1111 Franklin Street, 12th Floor, Oakland, California 94607 (US).

(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, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

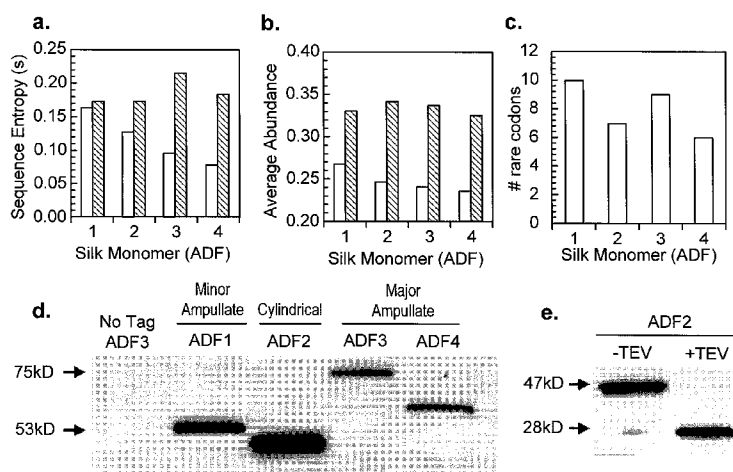
(72) Inventors; and

(75) Inventors/Applicants (for US only): VOIGT, Christopher Ashby [US/US]; 1855 Alabama Street #3, San Francisco, California 94110 (US). WIDMAIER, Daniel Matthew [US/US]; 344A Prentiss Street, San Francisco, California 94110 (US).

Published:

— without international search report and to be republished upon receipt of that report

(54) Title: BIOPOLYMER AND PROTEIN PRODUCTION USING TYPE III SECRETION SYSTEMS OF GRAM NEGATIVE BACTERIA



(57) Abstract: Proteins, polynucleotide, expression cassette, vector and bacterium compositions for obtaining proteins of interest by expression of same in gram negative bacteria having a Type III secretion system are provided. Uses of the proteins so obtained in the manufacture of isolated proteins, silks, fibers threads, and pharmaceutical compositions are also provided.

BIOPOLYMER AND PROTEIN PRODUCTION USING TYPE III SECRETION SYSTEMS OF GRAM NEGATIVE BACTERIA

CROSS-REFERENCES TO RELATED APPLICATIONS

5 [0001] NOT APPLICABLE

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] NOT APPLICABLE

10

REFERENCE TO A "SEQUENCE LISTING," A TABLE, OR A COMPUTER PROGRAM LISTING APPENDIX SUBMITTED ON A COMPACT DISK.

[0003] NOT APPLICABLE

15

BACKGROUND OF THE INVENTION

[0004] Proteins have a variety of biological, therapeutic, commercial and industrial uses. Many potential high-value proteins are difficult to harvest from their natural source and their production in microorganisms has proven difficult due to cloning, expression, and toxicity issues. In particular, natural biopolymers have evolved remarkable structural properties that are ideal materials for fabrics, plastics, glues and gums, and medical devices. For example, the family of spider silks include threads that are stronger than steel while maintaining flexibility and elastic properties. Human-derived biopolymers, such as elastin, are non-antigenic materials that can be used in medical devices. Many potential high-value biopolymers are particularly difficult to harvest from their natural source and their production in microorganisms has proven difficult due to cloning, expression, and toxicity issues.

20

25

[0005] Silks are protein-based biopolymer threads secreted by insects. They have evolved a remarkably wide range of desirable material properties. For example, the dragline threads that form the structural core of webs are stronger than Kevlar with ten times the elasticity. Changes in the amino acid sequence vary the flexibility, elasticity, strength, and stickiness of the threads.

30

[0006] Many natural silks have materials, industrial, and medical applications. However, with few exceptions, it is impossible to farm and extract the source of the material. Spiders produce very small quantities of silk and they cannot be raised in high densities due to their territorial behavior. Thus, there has been much effort to use recombinant DNA to produce these materials in high yield.

[0007] Gram negative bacteria use type III secretion (TTSS) to translocate proteins from the cytoplasm, through both membranes, to the extracellular environment (Galan and Collmer, 1999). A TTSS forms the core of several complex molecular machines, including the flagellum for swimming and molecular syringes that deliver effector proteins into plant and animal cells to promote pathogenesis and symbiosis (Huek, 1998; Chilcott and Hughes, 2000; Dale et al., 2002). While these systems are evolutionarily related, there are significant differences between their regulation, genomic organization, and supermolecular structure (Figure 3) (Aldridge and Hughes, 2002).

[0008] Gram negative bacteria use type III secretion to translocate proteins from the cytoplasm, through both membranes, to the extracellular environment (Galan and Collmer, 1999). *Salmonella* is a well-equipped intracellular pathogen that contains two type III secretion systems and effector proteins that are injected into the host cell to facilitate invasion, disrupt the host cytoskeleton, direct organelle trafficking, alter host gene expression, and regulate apoptosis (Galan and Zhou, 2000; Haraga et al., 2003; Shotland et al., 2003; Hernandez et al., 2004). These systems are encoded in two regions of genome, referred to as pathogenicity islands. *Salmonella* Pathogenicity Island 1 (SPI-1) is required for bacteria to invade epithelial cells (Lee and Falkow, 1992). SPI-2 is expressed in the intracellular environment of macrophages (Cirillo et al., 1998).

[0009] Heretofore, there has been no practical success in expressing full-length silk in a yeast or bacterium. The only time the spider drag line silk protein ADF-3 has been expressed full-length is in certain insect cells. Partial genes have been expressed in mammalian cells and plants (potato, Arabidopsis, tobacco) and goat milk. Experience has shown that the genes are unusually difficult to work with in bacterial systems. Recombination frequently knocks out portions of the gene. Because of skewed amino acid usage, insect codons do not work well in bacteria. This difference causes protein truncations to be translated because the ribosome pauses and falls off the mRNA in bacteria.

[0010] There has been much effort to use recombinant DNA technologies to produce silk-like proteins in high yield. However, the repetitive and self-assembly properties of natural biopolymers have made them difficult to passage, express, and recover in bacteria and yeast. DNA repetitiveness subjects the genes to frequent homologous recombination events

5 (Arcidiacono et al., 1998). Repetitive DNA with high GC content can also produce undesirable mRNA secondary structures, which can hinder expression (Arcidiacono et al., 1998). Finally, unusual codon usage causes ribosome pausing, resulting in protein truncations (Fahnstock and Irwin, 1997; Arcidiacono et al., 1998). The codon optimization of human tropoelastin was demonstrated to dramatically improve expression in *E. coli* (Martin et al., 10 1995). Even when the silk monomers express, they often self-assemble into threads or inclusion bodies that are difficult to redissolve (Huemmerich et al., 2004b). It has been observed that it is more difficult to spin high-quality threads from monomers extracted from inclusion bodies.

[0011] As a result of these problems, many large silk proteins having a region of 200 amino 15 acids or more with exact sequence identity cannot be practically expressed in microbes at all. Thus, most examples of fibroin expression in microbes involve artificial genes based on a repetitive consensus motif of 4 to 100 amino acids. (Prince et al., 1995; Lewis et al., 1996; Fahnstock et al., 1997b; McMillan et al., 1999; Huemmerich et al., 2004a; Bochicchio et al., 2005; Bandiera, et al., 2005). The genes are synthesized by first making a codon 20 optimized oligonucleotide of the repeat unit, which is then ligated to form a gene of a defined length. There are two problems with these constructs. First, the materials properties of natural silks rely on the non-repetitive portions of the sequence, especially variability in the glycine-rich elastic regions (van Beek et al., 2002). The non-repetitive are also important for nucleating the self-assembly of the fiber and maintaining solubility (Bini et al., 2004; 25 Huemmerich et al., 2004a). Second, the artificial genes still suffer from recombination events due to DNA repetitiveness (Fahnstock et al., 1997a). Silks based on artificial consensus repeat units often express well in yeast and bacteria. For example, artificial biopolymers based on the *Nephila clavipes* dragline sequence have been expressed at rates of 100 mg/L/day (Fahnstock et al., 1997a).

30 [0012] Besides microbes, many expression systems have been proposed. Expression systems based on potatoes and tobacco (Scheller et al., 2001), arabidopsis (Yang et al., 2005), and goat milk have all been used. The only example of ADF-3 and ADF-4 protein expression for proteins having a region of 200 or more amino acids with exact sequence

identity is in insect Sf9 cells (Huemmerich et al., 2004b). However, the biopolymers were not secreted and formed visible fibrils in the cytosol. A partial gene has been expressed in mammalian cells and secreted into the culture media (Lazarias et al., 2002). All of these expression systems are less economically viable than a microbial fermentation.

5 [0013] U.S. Patent Application Publication No. 20060068469 discloses the use of bacterial systems having a Type III secretion system to deliver proteins or polynucleotides to cells and animals. However, these systems rely upon the ability of the bacterium to enter a target cell or of the bacterium to lyse in the presence of a target cell.

[0014] Accordingly, this invention overcomes these difficulties by providing methods of
10 making proteins by re-engineering gram negative bacteria having a Type III secretion system.

BRIEF SUMMARY OF THE INVENTION

[0015] In a first aspect the invention provides an expression system based upon the Type III secretion system (TTSS) of gram negative bacteria. The expression system comprises genetic control elements which link the expression of a heterologous gene to the completion
15 of a functional TTSS in a gram negative bacteria. The expression system can utilize a natural genetic circuit or recombinant/hybrid genetic circuit to link the heterologous gene expression to the completion of a functional TTSS. In some embodiments, these genetic circuits can serve to control when the TTSS is expressed as well as the order and magnitude of the expression of one or more heterologous genes. The heterologous gene encodes a fusion
20 protein comprising an N-terminal polypeptide tag and a protein of interest. In some embodiments, the protein of interest is not naturally occurring in bacteria. The polypeptide tag serves as N-terminal secretion signal routing the peptide to the TTSS and can have the amino acid sequence of a native secretion signal tag or domain. Preferably, a tag is fused or attached to the protein of interest by an amino acid sequence which is subject to cleavage by
25 a protease to be subsequently contacted with the fusion protein once it is secreted into the medium. In preferred embodiments, the chaperone utilized by the TTSS is co-expressed with the tagged heterologous protein. In some further such embodiments, the chaperone utilized by the TTSS is co-expressed with the tagged heterologous protein under the control of a sicA promoter.

30 [0016] In a second aspect the invention provides a gram negative bacterium expressing a heterologous protein which is capable of being secreted into a growth medium via the TTSS.

In some embodiments, the bacterium has an expression system comprising genetic control elements which link the expression of a heterologous gene to the completion of a functional TTSS in a gram negative bacteria. In some embodiments, the bacterium is *Salmonella*, *E.Coli*, *Yersinia*, *Shigella*, and *Bordetella*. In some further embodiments, the bacterium is *Salmonella typhimurium*, and in still further embodiments, *Salmonella typhimurium* 1344. In some embodiments, the bacterium may be engineered to also express a heterologous protease which may, additionally, be engineered for secretion via the TTSS as a fusion protein. The bacterium may be modified so as to have a reduced or no ability to express an effector protein secreted by the secretion system. For example, multiple genes encoding effector proteins involved in virulence may be deleted or silenced. In addition, the bacteria may have been modified to eliminate the dependence of TTSS secretion on the presence of a host cell. Preferably, the contact dependence genes of the bacteria can for instance be knocked out or modified to remove the contact dependence (requirement for a host cell) to increase secretion into the media. In some embodiments, the bacterium is incapable of infecting natural host cells. In preferred embodiments, the bacteria is in a growth medium free of mammalian cells. In some embodiments, the medium is free of cells which can be entered by the bacterium. In some embodiments, the medium comprises cells which consist of, or consist essentially of, the gram negative bacterium. The heterologous protein comprises an N-terminal polypeptide tag and a protein of interest. The polypeptide tag serves as N-terminal secretion signal routing the peptide to the TTSS and can have the amino acid sequence of the native secretion signal tag. Preferably, the tag is fused or attached to the protein of interest by an amino acid sequence which is subject to cleavage by a protease to be subsequently contacted with the fusion protein once it is secreted into the medium.

[0017] In a third aspect, the invention provides a method of making a protein of interest or a fusion protein comprising the protein of interest, by expressing the protein in a gram negative bacterium having a type III secretion system and an expression system as described herein. The protein is heterologous to the bacterium and is expressed as a fusion protein comprising the protein of interest and an N-terminal polypeptide sequence or tag that directs the fusion protein to the TTSS which secretes the fusion protein into the medium of the bacterium. The fusion protein preferably comprises an amino acid sequence linking the tag and protein in which the amino acid sequence is subject to hydrolysis by a predetermined protease to be contacted with the fusion protein. The secreted fusion protein may be isolated from the medium prior to contacting the secreted protein with the protease. The secreted

protein may be contacted with the protease in the medium. The protein of interest may be isolated after it is released by contacting with the protease. In some embodiments, the protein of interest is obtained in a form which is 90%, 95%, 99%, or 99.9% free by weight percent of other proteins. In some embodiments, the protein is a therapeutic protein, enzyme, growth factor, hormone, cytokine, an antibody, or a fibroin. In some further embodiments, the protein of interest is human or mammalian. One advantage of our system is that the protein is secreted into the media and, hence, is much less likely to form "inclusion bodies."

Inclusion bodies form in cells when protein accumulates in large quantities (like bubbles). This is toxic and it is difficult to extract the protein afterward because the inclusion bodies will not dissolve. In some of any of the above embodiments, the protein is not a protein which specifically binds a polynucleotide.

[0018] In a fourth aspect, the invention provides methods of making pharmaceutical compositions comprising a therapeutic protein or a therapeutic fusion protein made by use of the above described methods, bacterium, or expression systems. The proteins may be formulated for any suitable method of administration according to the target site for the administration and effect sought. In some further embodiments, the therapeutic protein comprises or is a human or mammalian protein.

[0019] In fifth aspect, the invention provides polynucleotides encoding a fusion protein which comprises a protein of interest heterologous to gram negative bacterium and a tag sequence of gram negative bacterium. The bacterium in some embodiments is *Salmonella*, *E. Coli*, *Yersinia*, *Shigella*, and *Bordetella*. In some further embodiments, the bacterium is *Salmonella typhimurium*, and in still further embodiments, *Salmonella typhimurium* 1344. In some embodiments, the protein of interest is a fibroin protein or a therapeutic protein. In some embodiments, the protein of interest is a therapeutic protein, enzyme, growth factor, hormone, cytokine, or an antibody. In some further embodiments, the protein of interest is human or mammalian. In some embodiments of any of the above embodiments, the tag and the protein of interest are joined by an linker sequence amino acid sequence which is optionally has a sequence subject to cleavage by a predetermined protease to be contacted *in vitro*. In some embodiments, the polynucleotides are used in the above described expression systems, bacterium, and methods of making proteins or pharmaceutical compositions.

[0020] In another aspect, the invention provides a vector or expression cassette comprising the above-described polynucleotides. The vector is capable of or suitable for transfecting or

transducing a bacterium. In some embodiments the bacterium is *Salmonella*, *E. Coli*, *Yersinia*, *Shigella*, *Salmonella*, and *Bordetella*. In some further embodiments, the bacterium is *Salmonella typhimurium*, and in still further embodiments, *Salmonella typhimurium* 1344. In some embodiments, the polynucleotide is operably linked to a gene regulatory element that coordinates the expression of the protein of interest encoded by the polynucleotide with a functioning Type III secretion system.

[0021] In another aspect the invention provides a kit comprising a gram negative bacterium having no or reduced expression of an effector secreted by a Type III secretory system of the bacterium and a vector comprising a gene regulatory element, wherein the element is capable of coordinating the expression of a polynucleotide operably linked to the element with a functional TTSS system. In some further embodiments, the bacterium is *Salmonella*, *Salmonella typhimurium*, or *Salmonella typhimurium* 1344. For instance, in some embodiments, the kit comprises a first container holding a *Salmonella* bacterium having no or reduced expression of an effector secreted by a Type III secretory system, and a second container holding a polynucleotide encoding a *sicA* promoter.

[0022] Preferably, in any of the above aspects and embodiments, the chaperone for the tagged heterologous protein is co-expressed with the tagged heterologous protein. In some further such embodiments, the chaperone utilized by the TTSS is co-expressed with the tagged heterologous protein under the control of a *sicA* promoter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] **Figure 1.** Amino acid sequence of some exemplary human polymers.

[0024] **Figure 2.** Amino acid sequence of some exemplary TTSS secretion signals.

[0025] **Figure 3.** Regulatory Control of SPI-1. Regulatory interactions are colored by the order that they are activated in inducing conditions (top). A circuit diagram for this network is shown (bottom).

[0026] **Figure 4.** A genetic circuit links effector expression to the completion of the TTSS. **(left)** Before the TTSS is function, the effectors SipB and SipA (dark circle) bind to the chaperone SicA. **(center)** When the effector is secreted, the chaperone is free to bind to the transcription factor InvF. This activates the transcription factor, which upregulates effectors

from the internal *sicA* promoter. **(right)** Upon upregulation, the SicP chaperone directs the SptP effector for secretion.

[0027] Figure 5. SPI-1 promoters are turned on following a temporal order. The top bars indicate the standard deviation in on-times from six experiments performed on different days.

5 **[0028] Figure 6.** When the input signal is removed via dilution, the *sicA* promoter demonstrates hysteresis. (A) Flow cytometry data for the *hilC* and *sicA* promoters are shown (from top to bottom: 40, 205, 280, 380 minutes). The *hilC* and *hilD* promoters do not show hysteretic effects (B), but *sicA* does (C).

[0029] Figure 7. SPI-1 promoters exhibit both graded and all-or-none induction dynamics. 10 The induction of SPI-1 promoters of the initial activator HilC (left), the *prg* operon encoding the TTSS inner membrane ring (center), and the *sicA* promoter controlling the expression of effectors (right) is shown as a function of OD. From top to bottom, the OD₆₀₀ is 0.4, 0.6, and 1.8.

[0030] Figure 8. An expression time course is shown for the wild-type (SL1344) and 15 *ΔinvE* *Salmonella* strains. Each lane is secreted protein, obtained by precipitating the supernatant. The *ΔinvE* strain secretes significantly more protein and significant expression and secretion is observed after eight hours. The positive control is a standard FLAG-tagged protein at a concentration corresponding to 3 mg/L.

[0031] Figure 9. Secretion of Spider Silk. A Western blot is shown of the ADF-3 20 monomer fused to a FLAG tag. Protein was recovered using a secretion assay as described previously (Lee and Galan, 2004).

[0032] Figure 10. Amino acid sequence of FLAG tag.

[0033] Figure 11. The *Salmonella* type III secretion system is harnessed to export spider silk proteins. A synthetic control system (right) is constructed on a plasmid, and receives 25 information from the natural regulatory network controlling TTSS self-assembly and function (left). TTSS self-assembly is initiated by three transcription factors (HilC/D/A), which integrate environmental information. In turn, they initiate a transcriptional cascade that regulates the timing of the transcription of multiple operons. An operon containing the chaperones and effectors (SicA) is controlled by a genetic circuit that becomes active once 30 the TTSS is constructed and functional. This genetic circuit is used to drive the expression of a chaperone (SicP) and silk protein fused to an N-terminal secretion signal (SptP), all of

which are encoded on the pCASP plasmid. After secretion, the signal sequence (SptP) is removed by the TEV protease.

[0034] Figure 12. The type III secretion system is assembled by an ordered transcriptional cascade with a genetic circuit linking the upregulation of secreted effectors to the completion of functional secretion needles. (a) The temporal ordering of the SPI-1 promoters is shown using *gfp* fusion reporters and flow cytometry. First, the transcriptional regulators *hilD* (Δ) and *hilC* (x) accumulate. These regulators turn on the *prgH* promoter (o), which controls structural genes in the needle complex. Once the TTSS is functional, the effectors are strongly upregulated from the *sicA* promoter (+). The average gated fluorescence of each promoter is normalized by the maximum and the data represents four experiments for each promoter performed on different days. (b) A genetic circuit couples the upregulation of effector protein expression to the completion of functional needles. The InvF transcription factor is only active when bound to SicA dimers. Prior to the completion of the needle, SicA is bound and sequestered by the SipB/C effectors. As type III secretion needles become functional, SipB/C are exported and SicA binds free InvF to stimulate transcription of *PsicA*. (c) Cytometry data is shown, demonstrating that the *sicA* promoter is tightly off (dotted curve) until after the TTSS is complete (continuous).

[0035] Figure 13. The type III secretion system can export heterologous proteins. The human DH domain is used to determine secretion efficiency and longevity. (a) The DH domain is secreted when fused to the N-terminal SptP secretion signal and the SicP chaperone is co-transcribed (+SicP DH SL1344). There is no change in secretion when the flagella master transcriptional regulators are knocked out, demonstrating that secretion is flagella independent (DH Δ FlhCD). The quantity of protein secreted is dramatically reduced when the SicP chaperone is not co-transcribed (wt SicP). Secretion is not detected when the DH domain is expressed from the P_{BAD} promoter lacking the N-terminal signal (No Tag, Induced DH), but it can be detected when the cells are lysed (No Tag, Lysis). (b) Once the TTSS is constructed, DH protein is continuously produced and secreted. Prior to construction (at 2 hrs), no protein is secreted. After 8 hours, 60 mg/L is detected. (c) A coomassie stained gel of concentrated protein (equivalent to ~1.5mL of supernatant) is shown for lysed cells, (wt lysis) after the secretion assay for wild-type *Salmonella*, (wt supernatant) and cells secreting the DH domain (DH supernatant). The pattern of other secreted proteins in the supernatant matches those previously reported. The identity of the DH band is confirmed by immunoprecipitation of the protein from supernatant (DH Purified), the second band of

higher molecular weight is the heavy chain of the FLAG antibody. Below the gel is a western blot showing the periplasmic protein MalE is detectable in the lysate but not in either the wild-type or DH supernatants. This indicates that lysis is not significantly contributing to the proteins isolated from the secretion assay.

5 [0036] **Figure 14.** The expression and secretion of four *Araneus diadematus* spider silk monomers is shown. Each silk gene was optimized to eliminate repetitive DNA and rare codons. (a) The sequence entropy (Methods) was calculated as a measure of diversity for an alignment of the repeat units. The optimized genes (black bars) have more sequence diversity than the wild-type genes (clear bars). (b) The change in codon usage is shown. The *E. coli* 10 codon abundances were obtained from the KEGG database and averaged over the entire gene. (c) Very rare codons (defined as abundances <0.13) were entirely eliminated from the sequences. (d) A secretion assay is shown for the optimized ADF-1, 2, 3, and 4 genes (Lanes 2-5). When the N-terminal secretion tag is removed from the sequence, no protein is detected in the supernatant (Lane 1) (b) TEV protease cleaves the SptP secretion tag from silk 15 proteins. ADF-2 prior to digestion (Lane 1) is reduced in size by 19kD when the SptP secretion tag is removed by TEV protease.

[0037] **Figure 15.** The expression plasmid for heterologous secretion (pCASP).

[0038] **Figure 16.** Silk yields and secretion efficiency as measured by quantitative western blot.

20 [0039] **Figure 17.** Comparison of secretion tags. Each lane shows a secretion assay for different SPI-1 tags fused to the DH protein. In each case (except InvJ), the corresponding chaperone was co-expressed from the *sicA* promoter. The SptP tag yielded the highest secretion yield.

[0040] **Figure 18.** The utility of the *sicA* promoter is shown. (left) An assay is performed 25 to quantify the formation of inclusion bodies in the cell. When the silk genes are expressed from an IPTG-inducible promoter (PTRC), the ADF-3 and ADF-4 monomers form tight inclusion bodies. The expression from the *sicA* promoter produces less protein in the form of inclusion bodies. (right) The silk proteins can also be toxic when expressed and not secreted. This data shows cell growth as a function of time for different constructs expressing the 30 ADF-1 silk monomer. When expressed from the IPTG-inducible PTRC promoter, cell growth is significantly retarded. The effect on cell growth is eliminated when expressed from the *sicA* promoter and allowed to secrete.

DETAILED DESCRIPTION OF THE INVENTION

[0041] The invention relates to the use and modification of TTSS secretion system in gram negative bacteria to manufacture proteins. We have characterized the control systems involved in TTSS and identified useful methods of using and modifying TTSS to produce proteins of interest which are secreted into their medium by the bacteria.

[0042] Accordingly, in one aspect the invention provides a method of making a protein of interest by expressing the protein in a gram negative bacterium having a type III secretion system wherein the protein is heterologous to the bacterium and the protein is fused to a polypeptide or tag that directs the protein to the secretion system so that the protein is secreted via the secretion system into the medium of the bacterium. Preferably, the bacterium is *Salmonella*, *E.Coli*, *Yersinia*, *Shigella*, *Chlamydia*, or *Bordetella*. More preferably, the bacterium is *Salmonella typhimurium*, and still more preferably, the bacterium is *Salmonella typhimurium* 1344. In some embodiments of the above, the bacterium has a reduced ability to express a native effector protein secreted by the secretion system. For instance, multiple genes encoding effector proteins involved in virulence have been deleted or silenced. In other embodiments of any of the above, the Type III secretion system is an SPI-1 secretion system. In still further embodiments of any of the above, the bacterial strain used had contact dependence genes which have been knocked out to remove the contact dependence

(requirement for a host cell) to increase secretion from the resulting strain into the media. The heterologous protein or protein of interest can be a therapeutic protein, including human proteins. In some further embodiments, the heterologous protein or protein of interest is isolated from the medium and formulated with a pharmaceutically acceptable carrier. In some embodiments, the heterologous protein is a spider silk protein, an ADF protein, a silk worm silk protein, or elastin. Preferably, in any of the embodiments, a tag is coupled to the protein by a sequence of amino acids which is cleavable by an enzyme and the secreted protein is contacted with the enzyme and cleaved by the enzyme. In some embodiments, the enzyme is TEV protease and the sequence comprises the TEV recognition sequence.

Additionally, it is contemplated that the bacterium can be preferably modified to limit or eliminate their pathogenicity.

[0043] In additional embodiments of the above, the protein of interest is encoded by a gene wherein the DNA was structurally stabilized and optimized for expression in bacteria through

codon optimization, mRNA minimization, and reduction of recombination frequency. In other embodiments, the the heterologous protein is expressed by a gene on a plasmid.

[0044] In yet other embodiments of the foregoing, the expression of the protein is controlled by a genetic circuit that links the activation of expression to the completion of the TTSS structure. For instance, the expression of the heterologous protein can be operably linked or under the control of the *sicA* promoter in *Salmonella*. In some embodiments, accordingly, the expression of the protein is controlled by a SPI-1 promoter (e.g., *hilC*, *hilD*, *hilA*, *invF*, *sopE*, and *prgH*).

[0045] In further embodiments of any of the above, the bacterium is *Salmonella* and the tag is a tag of a *Salmonella* effector and the tag is an SptP effector tag (e.g., a tag comprising the sequence of SEQ ID No:1 or SEQ ID No:2).

[0046] In yet other embodiments of any of the above, microbial or *Salmonella* codons can be substituted for insect codons in the gene expressing the protein. In additional embodiments, the the repeat regions and mRNA structures are reduced by making use of the codon degeneracies.

[0047] In some embodiments of any of the above, the heterologous protein or protein of interest is a protein that can modify a chemical/biopolymer/substrate that cannot pass through the cellular membrane of the bacterium. In other embodiments, the expression of the heterologous protein (e.g., cellulase) is controlled by a stationary phase promoter (e.g., *spvA*, *spvR*, and *ssaG*).

[0048] In yet other embodiments of any of the above, the tag is the tag of the *SopE*, *InvJ*, or *SipA* effectors.

[0049] In still further embodiments of any of the above, other or competing secretion systems of the bacterium are knocked out. For instance, the SPI-2 and flagella can be knocked out or inactive. In additional embodiments of any of the above, one or more extracellular or intracellular proteases of the bacterium are knocked out or inactive.

[0050] In other embodiments of the foregoing, the expression of the protein of interest is under the control of a constitutive promoter or an inducible promoter (e.g., *lac*, *tet*, *PBAD*, etc.) as known to one of ordinary skill in the art. In addition, hybrid promoters that are IPTG inducible can be used to control the expression. For instance, *lac* operator sites can be located on either end of the promoter (e.g., *sicA*, *spvA*, or *ssaG*).

[0051] In yet other embodiments still, a *sicA* promoter drives the expression of T7 polymerase which then very strongly upregulates a T7 promoter.

[0052] In other embodiments of the above, the expression system comprises one or more genetic control element(s) which link the expression of a heterologous gene to the completion of a functional TTSS in a gram negative bacteria.

[0053] The invention also provides fusion proteins of a protein of interest and a polypeptide tag in which the protein is heterologous to wild-type *Salmonella*, and the tag directs a protein to a *Salmonella* Type III secretion system. In some embodiments, the protein comprises a spider silk protein; a silk worm silk protein; elastin; a fibroin; or a protein-based biopolymer; a blood or plasma protein; a mammalian peptide hormone, growth factor, cytokine, antibody, enzyme, or receptor. In some further embodiments, the tag is linked to a protein of interest via a amino acid sequence subject to hydrolysis by a predetermined protease which is specific to the amino acid sequence insofar as it does not internally cleave the protein of interest. In other embodiments, the protein further comprises a polypeptide tag or label used in the detection or purification of the protein.

[0054] In some embodiments of any of the above, preferably, where the bacterium has contact dependence genes which regulate the secretion system, the contact dependence genes are knocked out to remove the contact dependence (requirement for a host cell) for increased secretion into the media. Accordingly, in some embodiments of any of the above, the bacterium is modified so as to not be able to enter a host cell. In some embodiments, the bacterium is modified so as to not lyse upon contact with a host cell. In other embodiments, the contact dependence genes are knocked out to remove the contact dependence

[0055] In some embodiments of any of the above, the bacterium has a reduced ability to express an effector protein (e.g., wild type effector proteins, pathogenic effector proteins) secreted by the secretion system. For example, multiple genes encoding effector proteins involved in virulence can be deleted or silenced. Preferably, where the bacterium is *Salmonella*, the Type III secretion system is an SPI-1 secretion system. In yet other embodiments, the bacterium is modified so as not to bind a host cell or interact with a host cell it normally infects.

[0056] Preferably, in some embodiments of the foregoing, the tag is coupled to the protein by a sequence of amino acids which is cleavable by an enzyme and the secreted protein is contacted with the enzyme and cleaved by the enzyme. For example, the protease can be

TEV protease and the sequence is the TEV recognition sequence. The consensus TEV recognition sequence is Glu-X-X-Tyr-X-Gln/Ser. Cleavage occurs between the conserved Gln and Ser residues. X can be various amino acyl residues but note that not all residues are tolerated. (see, Dougherty et al. 1989. Virology 171:356-364).

- 5 [0057] Optimally, it is contemplated that the polynucleotide or gene encoding the protein of interest can be structurally stabilized and optimized for expression in bacteria through codon optimization, mRNA minimization, and reduction of recombination frequency. For instance, the repeat regions and mRNA structures of fibroins can be reduced by making use of the codon degeneracies.
- 10 [0058] In addition, it is also contemplated that the heterologous protein or protein of interest can be expressed by a gene on a plasmid in the bacterium. Additionally, in such embodiments, the expression of the protein is preferably controlled by a genetic circuit that links the activation of expression to the completion of the TTSS structure. Accordingly, in some embodiments of the invention in any aspect, the heterologous protein is expressed
- 15 under the control of the *sicA* promoter or a SPI-1 promoter. Suitable promoters include, but are not limited to, *hilC*, *hilD*, *hilA*, *invF*, *sopE*, and *prgH*. In still other embodiments, the expression of the protein is controlled by a stationary phase promoter. Suitable stationary phase promoters are also contemplated in some embodiments, These promoters include, but are not limited to, *spvA*, *spvR*, and *ssaG*.
- 20 [0059] In yet other embodiments, the expression of the protein is under the control of a hybrid promoter system that is IPTG inducible. For instance, promoters (e.g., *sicA*, *spvA*, *ssaG*) are located between lac operator sites (i.e., at either end of a promoter). The regulatory protein LacI binds to two sites and causes the promoter to loop, which shuts it off. In the presence of IPTG, the promoter opens up and the promoters are active (only if the bacterium
- 25 is in the correct growth/secretion state).
- [0060] In other preferred embodiments of any of the foregoing, competing secretion systems are either knocked out, not activated, or not expressed. These knocked out systems include, but are not limited to, including SPI-2 and flagella. In another set of embodiments of the foregoing, extracellular and/or intracellular proteases are knocked out, not activated, or
- 30 not expressed. Combinations of such knockouts (e.g., competing secretion systems, effector knockouts, and protease knockouts) provide a secretion/expression optimized strain of the bacterium (e.g., *Salmonella*, *E.Coli*, *Yersinia*, *Shigella*, *Chlamydia*, and *Bordetella*).

[0061] In another set of embodiments of any of the above, the control circuit serves to amplify the *sicA* promoter to increase expression/secretion. In some such embodiments, for instance, the *sicA* promoter drives the expression of T7 polymerase, which then very strongly upregulates a T7 promoter. This strategy allows greater expression, while still only turning on when the TTSS is complete and functional.

[0062] In some embodiments, the heterologous protein is introduced on a plasmid in which expression is controlled by a genetic circuit that links the activation of expression to the completion of the TTSS structure. For instance, specifically, this can be done by putting the gene under the control of the *sicA* promoter. Via a chaperone feedback mechanism, this promoter is only turned on when the TTSS is built and is functional. This avoids the expression and accumulation of potentially toxic proteins before the secretion system is ready. We have found *sicA* to be a remarkably strong promoter; much stronger than many common inducible systems.

[0063] In the case of *Salmonella*, we find that the SPI-1 system is particularly advantageous over other secretion systems in *Salmonella* because its induction was easily controlled in culture. It can be induced uniformly through the cells in standard LB media. To avoid the expression of SPI-1, the bacteria can be grown in other standard medias, such as M9 minimal media and Rich Broth, where SPI-1 is not induced. Accordingly, in some embodiments, the culture conditions used to culture the bacteria can be selected to avoid expression and bypass the load imposed by massive overexpression, until desired. Additionally, generally, the SPI-1 can also be induced artificially by using control elements which allow the HilC and HilD transcriptional activators to be expressed under conditions which otherwise would be non-inducing for SPI-1 or in a media where the needle would not otherwise be naturally activated.

[0064] In some further embodiments of the above, the bacterium is *Salmonella* and the tag is a tag of a *Salmonella* effector. In other embodiments, the tag is an SptP effector tag. In still further embodiments, the tag comprises the sequence of SEQ ID No:1, SEQ ID No:2. or SEQ ID NO:3 or one substantially identical thereto. Suitable tags include those of the SopE, InvJ, or SipA effectors.

[0065] In some embodiments if any of the above, *hilD* is placed under inducible control. The initiation of SPI-1 is controlled by the *HilD* transcriptional activator. This also is the primary activator of the *prg* operon, which controls the number of needles that are

constructed (Kubori et al., 2000). Secretion can be increased by overexpressing the HilD transcriptional activator from an inducible plasmid.

[0066] In yet other embodiments, the expression of the protein is optimized by substituting *Salmonella* codons for insect codons in the gene expressing the protein.

5 [0067] In some embodiments, a plurality of proteins of interest can be expressed in one bacterium using the above systems. The proteins can be encoded on genes of the same polynucleotide, vector, or plasmid. They can be subject to the control of the same gene regulatory elements. In further embodiments, the proteins are capable of being spun together into a fiber. In some embodiments, they are silk or fibroin proteins are each from the same or
10 different species. In some embodiments, there are a plurality of any silk or fibroin proteins described herein.

[0068] In some embodiments, bioreactor fermentations are adjusted to maintain SPI-1 optimal conditions over a longer period of time. In some embodiments, tags are based upon the SptP secretion signal. In other embodiments, secretion signals corresponding or having
15 the sequence of SipA, SopE, and InvJ are contemplated.

[0069] The invention also provides 1) polynucleotides encoding the above described protein(s) and/or expression control systems and 2) gram negative bacterium (e.g., *Salmonella*, *E. Coli*, *Yersinia*, *Shigella*, *Chlamydia*, and *Bordetella*), transfected with the polynucleotides. These bacterium can be modified to provide no or reduced expression of an
20 effector secreted by the Type III secretory system used to secrete the protein in the bacterium.

[0070] The invention also provides vectors or expression cassettes comprising the above-described polynucleotides. In some embodiments, the vector or expression cassette has the polynucleotide operably linked to a gene regulatory element that coordinates the expression of the protein encoded by the polynucleotide with a functioning Type III secretion system
25 (e.g., the *Salmonella* SPI-1 TTSS).

[0071] The invention also provides kits. The kit can comprise a *Salmonella* bacterium or other bacterium (*Salmonella*, *E. Coli*, *Yersinia*, *Shigella*, *Chlamydia*, and *Bordetella*) having no or reduced expression of an effector secreted by a Type III secretory system and a vector or expression cassette comprising a gene regulatory element capable of coordinating the
30 expression of a polynucleotide with the expression of a functioning Type III secretory system in the bacterium. The bacterium are preferably modified to limit or eliminate their

pathogenicity. In some embodiments, the invention provides a kit having a first container holding a *Salmonella* bacterium having no or reduced expression of an effector secreted by a Type III secretory system, and a second container holding a polynucleotide encoding a *sicA* promoter. In some further embodiments, the bacterium is *Salmonella*, *Salmonella*

5 *typhimurium*, or *Salmonella typhimurium* 1344.

[0072] In some further embodiments of any of the above aspects and embodiments, the strain is *Salmonella typhimurium* 1344 which is not a lab strain. In some embodiments, the strain is a knocked out strain which knock-outs serve to increase secretion, expression, and/or eliminate virulence. Essentially, *Salmonella typhimurium* 1344 SP-I TTSS can be

10 turned into a secretion-competent expression strain analogous to *E.Coli* BL21.

[0073] In some embodiments, the protein or fusion protein according to the invention is labeled. The label can be used to detect or facilitate isolation of the compound as known to one of ordinary skill in the art. Preferably, the label is part of a fusion protein and encoded by the polynucleotide encoding the secretion amino acid tag and the protein of interest. A

15 "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, chemical, or other physical means. For example, useful labels include ³²P, fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins which can be made detectable, e.g., by incorporating a radiolabel into the peptide or used to detect or be detected by antibodies

20 specifically reactive with the peptide. The label can be used to detect or facilitate isolation of the compound. The label can be linked to the protein or fusion protein by an amino acid sequence which is subject to hydrolysis by a protease. The amino acid sequence can be subject to hydrolysis by the same enzyme used to cleave the link between the secretion tag sequence and the protein of interest (e.g., the amino acid sequences to be cleaved can be the

25 same or susceptible to hydrolysis by the same enzyme, optionally performed in the same step).

[0074] With regard to *Salmonella*, four classes of genes make up the wild-type SPI-1 TTSS: structural, effector, chaperone, and regulatory. Effectors are directed to the needle structure by chaperones, which bind to an N-terminal secretion signal (Lee and Galan, 2004).

30 Proteins are probably passed through the 30 angstrom pore of the needle in a partially non-globular state (Stebbins and Galan, 2003). The chaperones have the capability maintain the effectors in a semi-unfolded state (Stebbins and Galan, 2001). They direct the effector to a

domain of the needle that can unfold the protein in an ATP-dependent manner, such that it can be exported (Akeda and Galan, 2005).

[0075] Regulatory genes are encoded in the *Salmonella* pathogenicity island that control the commitment to needle production and internally regulate the gene expression order.

5 Normally, the SPI-1 regulatory network is centered around a four-tiered cascade (Figure 3). There is an initiation circuit (centered on HilC/D/A) that integrates many environmental signals and commits to needle formation (Lucas and Lee, 2000). Post-commitment, the cascade controls the order in which different operons are expressed. This order is required for the proper assembly of the needle (Sukhan *et al.*, 2001).

10 [0076] At the end of the *Salmonella* cascade, a TTSS genetic circuit links the completion of functional needles with the upregulation of effector expression (Figure 4) (Darwin and Miller, 2001). The core of the circuit is formed by an interaction between a chaperone (SicA), effectors (SipB/C), and a transcription factor (InvF). Before the needle is formed, all three accumulate in the cytoplasm, but the chaperone is titrated out by an overabundance of
15 effector. However, once the needle is complete, the effector is released, which frees the chaperone to bind to the transcription factor. Only in this bound state can the transcription factor activate promoters to upregulate effector expression. We show that one can harness this bacterial control circuit as part of modified TTSS expression system to ensure that the heterologous proteins are only expressed when the system is capable of secreting them out of
20 the cell.

[0077] The *Salmonella* SPI-1 TTSS forms a needle-like structure that projects from the cytoplasm, through the inner and outer membrane, and extends 50 nanometers from the cell surface. Under fully activating conditions, there are several hundred needles per cell.

Effectors are directed to the needle structure by chaperones, which bind to an N-terminal secretion signal. Proteins are probably passed through the needle in a partially non-globular state (Stebbins and Galan, 2003). The chaperones have the capability maintain the effectors
25 in a semi-unfolded state and do not require nucleotide hydrolysis to function (Stebbins and Galan, 2001) and the needle pore is ~30 angstroms wide, which is large enough to allow small folded proteins to pass. All of the necessary structural proteins, chaperones, and most
30 of the effectors exist as a single contiguous island in the genome. Accordingly, in some further embodiments of any of the above aspects, the heterologous protein is a small folded protein.

[0078] The *Salmonella* pathogenicity island also contains genes that encode transcription factors that internally regulate the order and conditions in which different genes are expressed. The SPI-1 regulatory network is centered around a four-tiered cascade. There is an initiation circuit (centered on HilC/D/A) that integrates many environmental signals and commits to SPI-1. Post-commitment, the cascade controls the order in which different operons are expressed. This order facilitates the proper assembly of the needle.

[0079] A genetic circuit links the completion of functional needles with the upregulation of effector expression in *Salmonella*. In preferred embodiments, we use this circuit as part of our expression system to ensure that the heterologous proteins are only expressed when the system is capable of secreting them out of the cell. The core of the circuit is formed by an interaction between a chaperone (SicA), an effector (SipB/C), and a transcription factor (InvF). Before the needle is formed, all three accumulate in the cytoplasm, but the chaperone is titrated out by an overabundance of effector. However, once the needle is complete, the effector is released, which frees the chaperone to bind to the transcription factor. Only in this bound state does the transcription factor activate promoters to upregulate effector expression.

[0080] As a proof-of-principle, we have gene optimized the full length ADF-3 gene and shown that it can be expressed and secreted in *Salmonella* (Figure 15). ADF-3 was chosen as a target because of its interesting mechanical properties and solubility. A previous attempt to express full length ADF-3 in microbes failed (Fahnestock et al., 2000). Silk threads can be mechanically spun from an aqueous solution of only ADF-3 (Lazarias et al., 2002).

Other Eubacterial Secretion Systems.

[0081] There are several secretion systems based in *E. coli* that enable proteins to be delivered out of the cytoplasm. The major secretion systems used in the laboratory include the Sec system and the twin arginine translocation (TaT) system (Pohlschröder et al., 2005), both of which deliver protein to the periplasm rather than the extracellular environment. The Sec system is well characterized and involves a short signal sequence that localizes pre-protein to the cytoplasmic membrane. The Sec translocon hydrolyzes ATP and translocates the unfolded pre-protein into the periplasmic space. The TaT system is used to translocate folded proteins across the bacterial membrane. Like the Sec system, TaT secretes proteins into the periplasmic space rather than the extracellular environment. Recently, the flagellar export system has been used to secreting heterologous proteins in quantities of 1-15 mg/L (Majander et al., 2005). There are several problems with the flagellar system. Only a small

number of flagella are produced by each cell (even when over expressed, only 10s are present). Second, its regulation is intended to build a structure, rather than for continuous secretion. It naturally shuts down secretion after the flagellum is constructed (Chilcott and Hughes, 2000).

5 Definitions

[0082] Unless otherwise stated, the following terms used in the specification and claims have the meanings given below.

[0083] It is noted here that as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise.

[0084] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer. Methods for obtaining (e.g., producing, isolating, purifying, synthesizing, and recombinantly manufacturing) polypeptides are well known to one of ordinary skill in the art.

[0085] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0086] Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical

Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0087] As to "conservatively modified variants" of amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

[0088] The following eight groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (*see, e.g., Creighton, Proteins* (1984)).

[0089] A "protein of interest" can be any protein, including but not limited to, protein drug, therapeutic protein, cytokine, enzyme, hormone, receptor, growth factor, fibroin with the proviso that the protein is heterologous to the wild-type of the bacterium used to express the protein or fusion protein thereof. The protein of interest preferably can be a naturally occurring human, mammalian, or insect protein or one which is substantially identical thereto, or a conservatively modified variant thereof. The protein can have the amino acid sequence of a naturally occurring human, mammalian or insect protein. The protein can be elastin, a fibroin or protein-based biopolymers, including spider silks. It can be a silk fiber protein, including spider silk proteins and silk worm silk proteins. The protein can be ADF-3. The silk proteins can be reengineered to increase their strength as known to one of ordinary skill in the art or to facilitate their expression in the bacterial system employed in their manufacture. The protein (e.g., cellulose) may be one which modify a chemical/biopolymer/substrate (e.g., cellulose) that cannot pass through a cellular membrane. The protein can be a protein component of blood or plasma. In addition, the protein can be multi-component silk biopolymers (silk worm, other moths, spiders), abductin (a strong glue-like biopolymer from mollusks), elastin and other human extracellular matrix proteins, or sericins and gum proteins (from insects). In addition a protein can be ADF-1, ADF-2, ADF-3

or ADF-4. ADF-3, for instance, is a dragline silk. ADF-1 and ADF-2 are flagelliform (the circles in a web) and ampullate (egg sack) silks. ADF-4 is a dragline silk with different properties than ADF-3. In some embodiments, the protein is an artificial protein-based biopolymer based on consensus repeat amino acid sequences. In other embodiments, the protein is a biopolymer or fibroin that contains heterologous functional domains inserted into their sequence (e.g., enzymes or functional groups that promote cell adhesion or other domain that is beneficial for a medical device). In some embodiments, the proteins of interest (e.g., fibroins, etc.) are capable of forming threads, fibers, films. The proteins may be able to self-assemble or to be spun by machinery into the threads or fiber as is known to one of ordinary skill in the art. Some suitable spider silks are described in Rising et al. Zoological Science, 22: 273-281 (2005). Drag line silk proteins are also suitable. Once obtained, silk proteins and fibroins and the like can be mechanically spun from monomeric solutions as known to one of ordinary skill in the art.

[0090] In some embodiments, the heterologous protein further is not a bacterial protein or is not an effector protein found in a wild-type bacteria. In some embodiments, the heterologous protein further is not a naturally occurring bacterial protein. In some embodiments, the heterologous protein further does not specifically bind nucleic acids. In some embodiments, the protein further is not an effector protein. In some embodiments in any aspect, the protein of interest can be at least 10, 20, 30, 40 or 50 amino acids long. For instance, the protein of interest can be from 10 to 50 amino acids long, 20 to 100 amino acids long, or 40 to 200 amino acids long, or 200 to 400 amino acids in length.

[0091] The "proteins of interest" can be obtained by hydrolysis of a fusion protein comprising the protein of interest by a protease specific for the amino acid sequence joining or linking the protein of interest to the tag sequence in the fusion protein. The proteins may be further isolated and purified as known to one of ordinary skill in the art. If intended for pharmaceutical uses, the proteins may be formulated with a pharmaceutically acceptable carrier.

[0092] A protein of interest can be a protein which is identical or substantially identical to a naturally occurring protein. A protein of interest may differ from a naturally occurring protein according one, two, three or more conservative amino acid substitutions.

[0093] Proteins of interest include therapeutic proteins. Therapeutic proteins are polypeptides which are administered to treat a disease or disorder in a subject. The subject

can be a human, primate, mammal, or any animal. The therapeutic proteins can be, for instance, an antibody (monoclonal or polyclonal, and/or humanized antibody), an enzyme, a hormone, a cytokine, a receptor, a ligand of a receptor, a clotting factor. The therapeutic protein can be identical to or substantially identical to any mammalian or human enzyme,
5 hormone, enzyme, cytokine, receptor, ligand for a receptor, or protein having a therapeutic use.

[0094] In particular, the protein of interest can be a protein that self-assembles into structure(s) with materials properties (fibers, threads, gums, films) important for industrial and medical applications. In preferred embodiments, the protein of interest is a natural
10 fibroin, elastin, or other macrobiopolymer.

[0095] Fibroins represent a large family of natural polymers that self-assemble from monomeric protein subunits. Many insects, especially moths and spiders, build silk threads as part of their webs, cocoons, and egg sacks. Silk worm silk is used as a common material because of high production rates and ease of farming. However, other insects and spiders
15 produce silks with varied and desirable properties, but do so in small amounts and many cannot be farmed.

[0096] Fibroins evolved a remarkably wide range of desirable material properties. For example, the dragline threads that form the structural core of webs are stronger than Kevlar with ten times the elasticity (Hinman et al., 2000). Flagelliform silk can be stretched three
20 times its length before breaking (Hayashi and Lewis, 2001). Changes in the amino acid sequence vary the flexibility, elasticity, strength, and stickiness of the threads (Gosline et al., 1999; Gatsey et al., 2001). Similar polymers are produced in higher animals, including humans, and are used to form the structural core of tissues and organs. The production of these proteins could be used as a non-antigenic material for medical devices. Recombinant
25 protein-based biomaterials can also be engineered to include functional domains that act as catalysts or guides for cell motility and tissue behavior (Maskarinec and Tirrell, 2005). Spider silk fibroins have also been shown to be useful in building artificial tissue (Tsubouchi et al., 2005; Dal Pra et al., 2005).

[0097] Spiders and insects have elaborate organelles that change the ambient conditions to
30 spin the silk threads (Vollrath and Knight, 2001). Natural silks tend to be composed of two or more monomeric variants (Dicko et al., 2004). The insect can vary the physical properties of the thread by changing the spinning conditions and the relative composition of monomers.

Purified monomers can self-assemble into threads when the pH of a solution is lowered (Huemmerich et al., 2004b). Long threads can be obtained using industrial polymer-spinning equipment). It has been shown that threads can be obtained from monomeric solutions of recombinant ADF-3 and ADF-4 (expressed in mammalian cells) that have similar properties of the natural silk (Lazarias et al., 2002).

[0098] The dragline thread of *Araneus diadematus* (common garden orbweaver) is primary comprised of two proteins: ADF-3 and ADF-4 (Huemmerich et al., 2004b). These biopolymers are made up of alternating alanine- and glycine- rich repeat sequences that give the thread its mechanical and elastic properties, respectively (Gosline et al., 1999). Dragline silk for the structural anchor of the web and has the strongest and physical properties, while maintaining elasticity (Gosline et al., 1999). In addition to the repetitive regions, there are non-repetitive domains at the N- and C- termini of the biopolymer. These domains are important for solubility (Lazarias et al., 2002) and may be involved in initiating the proper self-assembly of the threads (Bini et al., 2004; Huemmerich et al., 2004a).

[0099] Proteins have many medical and material uses, but they are too expensive to produce in bulk. This system is suitable for production of a variety of such biopolymers, including silkworm silk and human elastin. Elastin is a biopolymer that would be useful in making medical devices because it is extremely strong and not immunogenic. This method is also generally advantageous as in, for instance, the large scale production of heterologous protein. These proteins include, but are not limited to, proteins that have to be folded to be functional.

[0100] "Nucleic acid" or "polynucleotide" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

[0101] Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically,

degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

[0102] Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes an individual antigen or a portion thereof) of a protein of interest or may comprise a variant of such a sequence as set forth above. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded chimeric protein is not diminished, relative to a chimeric protein comprising native antigens. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90%, 95%, 96%, 97%, 98% or 99% identity to a polynucleotide sequence that encodes a native polypeptide or a portion thereof.

[0103] Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

[0104] Methods of non-viral delivery of nucleic acids include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, *and* agent-enhanced uptake of DNA. Lipofection is described in, *e.g.*, U.S. Patent No. 5,049,386, U.S. Patent No. 4,946,787; and U.S. Patent No. 4,897,355) and lipofection reagents are sold commercially (*e.g.*, TransfectamTM and LipofectinTM). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (*ex vivo* administration) or target tissues (*in vivo* administration).

[0105] An “expression cassette” refers to a polynucleotide molecule comprising expression control sequences operatively linked to coding sequence(s).

[0106] A “vector” is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

5 [0107] “Control sequence” or “control element” refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and terminators; in eukaryotes, generally, such control sequences include promoters, terminators and, in some
10 instances, enhancers. The term “control sequences” is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

[0108] “Operably linked” refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence
15 “operably linked” to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

[0109] The TTSS genes are generally expressed over a remarkably wide range of growth phases (OD_{600} 0.1-2.2). The natural SPI-1 promoters can be combined with synthetic genetic circuits to create a regulatory ‘assembly line,’ where different genes are programmed to be
20 expressed at different times. This approach is advantageous where biopolymer-modifying proteins are secreted at different stages of growth. First, after a protein is expressed and secreted, a protease can be secreted to automatically cleave the N-terminal secretion signal. Second, the TTSS system can be engineered to secrete a cellulase to convert cellulose, which does not cross the cell membrane, to glucose, which can be used for growth and as a
25 metabolic precursor. This enables the conversion biomass into drugs, specialty chemicals, and energetic compounds.

[0110] Nucleic acid is “operably linked” when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates
30 in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally,

"operably linked" means that the DNA sequences being linked are near each other, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

[0111] Each polynucleotide or gene can be codon optimized for expression in eubacteria and minimized for mRNA secondary structure. The degeneracy of the codon code can be used to reduce the repetitiveness of the DNA to avoid homogenous recombination. The performance of each gene can be tested for secretion in the expression system.

[0112] "Conservatively modified variants" also applies to nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence with respect to the expression product, but not with respect to actual probe sequences.

[0113] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, including siRNA and polypeptides, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a

comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection (*see, e.g.*, NCBI web site

<http://www.ncbi.nlm.nih.gov/BLAST/> or the like). Such sequences are then said to be

5 “substantially identical.” This definition also refers to, or may be applied to, the complement of a test sequence. The definition also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described below, the preferred algorithms can account for gaps and the like. Preferably, identity exists over a region that is at least about 25 amino acids or nucleotides in length, or more preferably over a region that is
10 50-100 amino acids or nucleotides in length.

[0114] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Preferably, default
15 program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0115] A “comparison window,” as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to the
20 full length of the reference sequence, usually about 25 to 100, or 50 to about 150, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art.

Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local
25 homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat’l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr.,
30 Madison, WI), or by manual alignment and visual inspection (*see, e.g.*, *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

[0116] A preferred example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are

described in Altschul *et al.*, *Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990), respectively. BLAST and BLAST 2.0 are used, with the

parameters described herein, to determine percent sequence identity for the nucleic acids and proteins of the invention. Software for performing BLAST analyses is publicly available

5 through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database

sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*).

10 These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated

using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino

15 acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the

quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of

either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences)

20 uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as

defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of

25 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

[0117] “Naturally-occurring” as applied to an object refers to the fact that the object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

30 **[0118]** The phrase “stringent hybridization conditions” refers to conditions under which a probe hybridizes to its target subsequence, typically in a complex mixture of nucleic acids, but to no other sequences. Stringent conditions are sequence-dependent and can be different

in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally,

5 stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium).

10 Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and
15 0.1% SDS at 65°C.

[0119] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize
20 under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency. Additional
25 guidelines for determining hybridization parameters are provided in numerous reference, e.g., and *Current Protocols in Molecular Biology*, ed. Ausubel, *et al.*, John Wiley & Sons.

[0120] For PCR, a temperature of about 36°C is typical for low stringency amplification, although annealing temperatures may vary between about 32°C and 48°C depending on primer length. For high stringency PCR amplification, a temperature of about 62°C is
30 typical, although high stringency annealing temperatures can range from about 50°C to about 65°C, depending on the primer length and specificity. Typical cycle conditions for both high and low stringency amplifications include a denaturation phase of 90°C - 95°C for 30 sec - 2 min., an annealing phase lasting 30 sec. - 2 min., and an extension phase of about 72°C for 1 -

2 min. Protocols and guidelines for low and high stringency amplification reactions are provided, e.g., in Innis *et al.* (1990) *PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc. N.Y.).

[0121] Knock-out cells and transgenic bacteria can be made by insertion of a marker gene or other heterologous gene into an endogenous gene site in the mouse genome via homologous recombination. Such mice can also be made by substituting an endogenous with a mutated version of the gene, or by mutating an endogenous, e.g., by exposure to carcinogens.

[0122] The term "heterologous" when used with reference to a protein or a nucleic acid indicates that the protein or the nucleic acid comprises two or more sequences or subsequences which are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid. For example, in one embodiment, the nucleic acid has a promoter from one gene arranged to direct the expression of a coding sequence from a different gene. Thus, with reference to the coding sequence, the promoter is heterologous. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein). "Heterologous" accordingly includes those proteins and polynucleotide sequences which are not found in a bacteria in which they are introduced. Such proteins can be of mammalian, primate, human, reptilian, or insect origin.

[0123] The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

[0124] The term "isolated" with regard to polypeptide or peptide fragment or polynucleotides as used herein refers to a polypeptide or a peptide fragment or polynucleotide which either has no naturally-occurring counterpart or has been separated or purified from components which naturally accompany it, e.g., in normal tissues such as lung, kidney, or placenta, tumor tissue such as colon cancer tissue, or body fluids such as blood, serum, or

urine. Typically, the polypeptide or peptide fragment or polynucleotide is considered "isolated" when it is at least 70%, by dry weight, free from the proteins and other naturally-occurring organic molecules with which it is naturally associated. Preferably, a preparation of a polypeptide (or peptide fragment thereof) or polynucleotide of the invention is at least 80%, more preferably at least 90% or 95%, and most preferably at least 99%, by dry weight, the polypeptide (or the peptide fragment thereof), or polynucleotide, respectively, of the invention. Thus, for example, a preparation of polypeptide x is at least 80%, more preferably at least 90%, and most preferably at least 99%, by dry weight, polypeptide x. Since a polypeptide or polynucleotide that is chemically synthesized is, by its nature, separated from the components that naturally accompany it, the synthetic polypeptide is "isolated."

[0125] An isolated polypeptide (or peptide fragment) or polynucleotide of the invention can be obtained, for example, by extraction from a natural source (e.g., from tissues or bodily fluids); by expression of a recombinant nucleic acid encoding the polypeptide; or by chemical synthesis. A polypeptide or polynucleotide that is produced in a cellular system different from the source from which it naturally originates is "isolated," because it will necessarily be free of components which naturally accompany it. The degree of isolation or purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis. In preferred embodiments, the fusion protein or protein of interest obtained or made according to the invention is isolated or purified into a state comprising at least 90, 95%, 98%, 99%, or 99.9% by weight of the protein as compared to any other proteins.

[0126] Once a recombinant chimeric protein is expressed, it can be identified by assays based on the physical or functional properties of the product, including radioactive labeling of the product followed by analysis by gel electrophoresis, radioimmunoassay, ELISA, bioassays, *etc.*

[0127] Once the encoded protein is identified, it may be isolated and purified by standard methods including chromatography (e.g., high performance liquid chromatography, ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. *See, generally, R. Scopes, Protein Purification*, Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology Vol. 182: Guide to Protein Purification*, Academic Press, Inc. N.Y. (1990). The actual conditions

used will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, *etc.*, – and can be apparent to those having skill in the art.

[0128] The invention also provides methods of making pharmaceutical compositions, wherein a protein or polypeptide made according to the methods of the invention is
5 formulated in a pharmaceutically-acceptable solution for administration to a cell or an animal, either alone or in combination with other components.

[0129] The protein so obtained can be administered directly to a subject as a pharmaceutical composition. Administration is by any of the routes normally used for introducing such a protein into ultimate contact with the tissue to be treated, preferably the
10 mucosal membrane and epithelial cells. The compositions comprising such proteins are administered in any suitable manner, preferably with pharmaceutically acceptable carriers. Suitable methods of administering such proteins are available and well known to those of skill in the art. Although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective
15 reaction than another route.

[0130] Pharmaceutical compositions comprising the proteins made or obtained according to the invention may be formulated in conventional manner using one or more physiologically acceptable carriers, diluents, excipients or auxiliaries which facilitate processing of the polypeptides into preparations which can be used pharmaceutically. Proper formulation is
20 dependent upon the route of administration chosen.

[0131] Pharmaceutically acceptable carriers, diluents, or excipients are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there are a wide variety of suitable formulations of pharmaceutical compositions of the present invention. For example, pharmaceutical
25 compositions can be formulated for topical administration, systemic formulations, injections, transmucosal administration, oral administration, inhalation/nasal administration, rectal or vaginal administrations. Suitable formulations for various administration methods are described in, *e.g.*, *Remington's Pharmaceutical Sciences*, 17th ed. 1985.

[0132] Briefly, for topical administration, the proteins made or obtained according to the
30 invention may be formulated as solutions, gels, ointments, creams, suspensions, *etc.* Systemic formulations include those designed for administration by injection, *e.g.* subcutaneous, intravenous, intramuscular, intrathecal or intraperitoneal injection, as well as

those designed for transdermal, transmucosal, oral or pulmonary administration. For injection, the proteins may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hank's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. For oral administration, a composition can be readily formulated by combining the proteins with pharmaceutically acceptable carriers to enable the chimeric proteins to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like. For administration by inhalation, the proteins obtained according to the present invention are conveniently delivered in the form of an aerosol spray from pressurized packs or a nebulizer, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. The proteins may also be formulated in rectal or vaginal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

[0133] Other suitable formulations and administration methods will be readily apparent to one of skill in the art and can be applied to the present invention.

[0134] After biological expression or purification, the heterologous proteins may possess a conformation substantially different than the native conformations of the constituent proteins. In this case, it can be helpful to denature and reduce the chimeric protein and then to cause the protein to re-fold into the preferred conformation. Methods of reducing and denaturing polypeptides and inducing re-folding are well known to those of skill in the art (*see* Debinski *et al.*, *J. Biol. Chem.* 268:14065-14070 (1993); Kreitman & Pastan, *Bioconjug. Chem.* 4:581-585 (1993); and Buchner *et al.*, *Anal. Biochem.* 205:263-270 (1992)). Debinski *et al.*, for example, describe the denaturation and reduction of inclusion body polypeptides in guanidine-DTE. The polypeptide is then refolded in a redox buffer containing oxidized glutathione and L-arginine.

Regulatory dynamics of type III secretion

[0135] Preliminary experiments were done to determine the suitability of the system-level dynamics of SPI-1 regulation. The transcription factor cascade was found to control the temporal order of gene activation, which reflects requirements for the self-assembly of the needle. As the effectors are regulated by a positive feedback loop, hysteresis was identified in which effector expression is slow to turn off when cells are shifted to non-inducing

conditions. Finally, the promoters were found to differ in their stochastic properties. Structural genes are turned on in an all-or-none manner whereas the effectors follow graded induction.

[0136] We have also shown that the SPI-1 system can be used to secrete heterologous proteins. An N-terminal secretion tag or signal was fused to the protein, which was placed under the control of a genetic circuit that links its expression to the completion of a functional TTSS. To improve expression, in some embodiments, one can use a knockout strain of *Salmonella* ($\Delta invE$), which makes the TTSS leak protein into the media in the absence of host cells (Zierler and Galan, 1995). For a well-expressed human protein, we obtained titers in the supernatant of ~20-40 mg/L. To express ADF-3 in bacteria, the gene was codon optimized while minimizing mRNA secondary structure and sequence repetitiveness. The optimized gene was chemically synthesized. This gene is shown to express and secrete at a titer of ~10-100ul/L.

[0137] Preliminary experiments have been performed to determine the suitability of the regulatory dynamics of the SPI-1 pathway. Reporter plasmids were constructed by transcriptionally fusing each SPI-1 promoter to green fluorescent protein. Strains containing the reporters were induced by growing to stationary phase in LB broth and the fluorescence was measured in single cells by flow cytometry. This approach was similar to that used to study the temporal regulation of flagella assembly (Kalir et al., 2001) and amino acid metabolism (Zaslaver et al., 2004).

[0138] From these experiments, three genetic circuits emerged as being particularly important in dictating the network dynamics. There is a genetic circuit early in the pathway that integrates two inputs (HilD/C/A) and commits the cell to the expression of structural and effector genes. Then, a four-tiered transcription factor cascade controls the temporal order of genes that are expressed. At the end of the pathway, there is a strong positive feedback loop (SicA and InvF) that links the completion of a functional TTSS with the upregulation of effectors.

[0139] We also discovered a unique ability for the network to differentially control the stochastic properties of promoters in the same cascade. Structural genes are controlled in an 'all-or-none' manner, whereas effectors are induced with graded dynamics.

[0140] We have also characterized the temporal ordering of gene expression. When SPI-1 is induced, there is a temporal order in which the operons are transcribed (Figure 4). As the

cell density increases, the HilD and HilC inputs are expressed first. Next, *hilA* promoter and the *prg* operon are activated, which encode the structural genes forming the inner membrane ring and needle (Kubori et al., 2000). Then, the *inv* genes are expressed, which contain structural genes that form the outer membrane ring and functional genes. The observed
5 temporal ordering is consistent with electron microscopy studies, which demonstrate that the inner membrane ring must form before the outer membrane ring (Sukhan et al., 2001). Finally, chaperones and effectors are expressed from the *sicA* and *sopE* promoters. Similar transcriptional ordering has been observed during the assembly of the evolutionarily related flagella basal body in *E. coli* (Kalir et al., 2001) and *Caulobacter crescentus* (Laub et al.,
10 2000).

[0141] The range of on-times is remarkably broad (OD₆₀₀ 0.1-1.2). In addition, there are late stage promoters that turn on at even higher OD's, up to about 2.2 (not shown). Accordingly, these promoters can be advantageously combined with synthetic regulation to create an 'assembly line' of protein expression.

15 [0142] We have also found evidence of hysteresis in effector expression. Effectors are controlled by a positive feedback loop, where SicA activates its own promoter (Figures 4 and 5). This feedback loop causes the *sicA* promoter to continue to be active after the cells are shifted to non-inducing conditions (Figure 6). We performed a simple dilution experiment, where aliquots are taken at the peak of SPI-1 induction, washed, and rediluted 1:100 into
20 fresh LB. After dilution, the fluorescence of the *hilD* and *hilC* reporters decay rapidly, whereas there is persistence in the expression from the *sicA* promoter (Figure 6). This feedback loop ensures that effector genes continue to be expressed even when the SPI-1 structures are complete and the structural genes are down regulated. Indeed, the effector promoters are active after 24 hours, when then the promoters controlling structural genes are
25 downregulated (not shown).

[0143] We have also characterized the *all-or-none* and *graded induction of TTSS Genes*. We have discovered that the SPI-1 promoters fall into three classes of stochastic behaviors. Depending on whether a promoter controls regulatory, structural, or effector genes, it is induced in a graded, all-or-none, or mixed manner (Figure 7). This stochastic control can
30 enable the cell to commit to the formation of a fixed number of TTSS needles and then increase the expression of effectors as a function of the environmental conditions. The *sicA* promoter, which forms the core of our synthetic inducible system, has both all-or-none and

graded properties. Most suitably, this finding indicates that the heterologous protein will continue to be expressed after the needle is complete.

[0144] Sequences can be optimized for expression in the eubacteria. For instance, the amino acid sequences of silks and fibroins contain highly repetitive regions. This often results in a high rate of homologous recombination, protein truncation, and mRNA secondary structure. Computational methods have been designed to overcome these problems. Such tools were applied to the full length ADF-3 dragline silk gene. There is little or no evidence of homologous recombination using this gene. This is evidenced by a single band of protein after even 24 hours of growth (Figure 9). Multiple bands have been shown to occur when recombination-sensitive proteins are expressed (Arcidiacono et al., 1998).

[0145] We have demonstrated heterologous expression can be achieved through the TTSS. We constructed a plasmid-based system to secrete heterologous proteins through the SPI-1 TTSS (Figure 4). This system fuses an N-terminal secretion signal to the heterologous protein. The gene is placed under the control of the *sicA* promoter such that it is only expressed when the TTSS has been constructed and is functional. A FLAG-tag is included at the C-terminal end of the protein so that Western Blots can be performed to detect secretion. It has been previously shown that the FLAG-tag will pass through SPI-1 (Ham et al., 1998; Lilic et al., 2006). Genes can easily be cloned into this vector and be tested for secretion efficiency.

[0146] The N-terminal secretion signal is taken from the SptP effector (Lee and Falkow, 2004). This tag interacts with the SicP chaperone, which has the capability of maintaining effectors in a partially unfolded state (Stebbins and Galan, 2001). The secretion signal is 167 amino acids, where residues 15-100 interact with a SicP dimer (Fu and Galan, 1998). It has been shown previously that a small peptide can be inserted into SptP and it will still be delivered through the TTSS to a host cell (Russman et al., 1997).

[0147] The secretion system was tested with two heterologous proteins. The first is a 24 kD folded human protein (DH domain of intersectin), which was shown previously to express well in bacteria (Rossman et al., 2002). A secretion assay is performed as described previously (Lee and Galan, 2004). Figure 8 shows a time course of DH secretion. Each point represents directly loaded supernatant; additional precipitation was not necessary. In the wild-type SL1344 strain, protein starts to accumulate 16 hours after inoculation and accumulates over 24 hours. To boost secretion, we obtained a $\Delta invE$ knockout. This has been shown

previously to eliminate the need for host cell contact and boost effector expression into the media (Kubori and Galan, 2002). Significantly more protein is obtained in this knockout, especially after 8 hours. Based on a standard curve (Sigma Aldridge FLAG-BAP fusion protein P7457), we estimate that protein expression is on the order of 20-40 mg/L.

5 [0148] With minimal strain optimization, we have obtained 20-40 mg/L of secreted protein in the *Salmonella* system. A single gene knockout ($\Delta invE$) yields significantly more protein (Figure 8).

[0149] The optimized ADF-3 gene also expresses and is secreted (Figure 9), albeit at lower concentrations (10-100 ug/L). It is noteworthy that only a single band occurs at the correct
10 weight, indicating that homologous recombination is not a problem. It is unclear if the lower yield is due to less expression or inefficient secretion. There are no obvious growth defects when ADF-3 is expressed (not shown).

Methods

15 [0150] This section contains the details of experimental methods used to generate the preliminary results. These methods can be used to collect secreted proteins and perform Western Blots, determine the temporal dynamics of synthetic promoters, and perform genetic knockouts.

[0151] *Secreted Protein Collection.* Acetic acid extractions have been shown previously to
20 be effective at collecting silk biopolymers (Mello et al., 2004)

[0152] *Western Blotting.* The western blotting for DH-FLAG and sADF-3-FLAG was carried out as follows. 5uL of filtered secretion supernatant is run out NEAT on a 10% polyacrylimide gel (Biorad # 161-1101) at a constant amperage of 35mA until the dye front
25 reaches the edge of the gel. Proteins are transferred to a PVDF membrane (Biorad # 162-0186) at a constant 18V for 45-50 minutes. This blot is blocked with 5% Non-fat milk (Biorad cat#170-6404) TBST (125mM NaCl, 25mM Tris-base, 0.05% TWEEN20, pH 7.4) for 1 hour to overnight at room temperature. Primary antibody (Mouse anti-FLAG, Sigma, # F3165) is diluted 1:10,000 in TBST with 5% non-fat milk. This solution is allowed to bind to
30 the membrane for 1 hour at room temperature. Membranes are washed three times under tap water followed by three 10 minute washes in TBST. Secondary antibody (HRP-Sheep anti-Mouse) is diluted 1:5,000 in TBST-5% non-fat milk and applied to the blot for 1 hour at room temperature. The blot is washed three times under tap water followed by three 10

minute washes in TBST. The blot is developed by applying 2mL of prepared developer reagent (Pierce Cat# 32209) to the blot on a clean piece of cling wrap for 1 minute. Excess developer is wicked away with Kimwipe and the blot is placed face down on a clean piece of cling film and air bubbles rolled out. The blot is taped into a development cassette and
5 images acquired using chemiluminescent film (Kodak #178-8207). Typical exposure times are 60 - 180 seconds.

Ways to Optimize the *Salmonella* strain for Maximum Heterologous Protein Expression and Secretion.

10 [0153] As noted above, with minimal strain optimization, we have obtained 20-40 mg/L of secreted protein. The wild-type strain produces about 2-fold less secreted protein than the Δ invE strain. The goal of this aim is to systematically optimize the secretion signal, strain, and promoter to maximize the titer of secreted protein. In addition, the total weight of effectors secreted by SPI-1 can be as high as several grams per liter. These effectors have not
15 yet been knocked out in our system and may compete with the synthesis and secretion of heterologous protein. Accordingly, in some embodiments, these competing effectors or wild-type effectors are knocked out. Only 50-70% of the cells in culture induce SPI-1, depending on the conditions (Figure 7).

[0154] In some embodiments, the bacteria are modified to increase the fraction of bacteria
20 expressing SPI-1 as well as the total number of needles using synthetic genetic circuits. In addition, SPI-1 expression reaches a peak after 16 hours and then decreases in very late stationary phase (not shown). This is due to our experimental design, where the cultures are grown in batch. Bioreactor fermentations will allow us to maintain SPI-1 optimal conditions over a longer period of time. The current system utilizes the SptP secretion signal and SicP
25 effector. This is the most studied signal-chaperone pair, but the SptP effector is not highly secreted into the culture media. Accordingly, other effector signals are contemplated (e.g., SipA, SopE, and InvJ are secreted to a concentration 100-fold higher (Lee and Galan, 2004). In some embodiments, the wild-type *sicA* promoter and ribosome binding site are used. Higher expression may be achieved by using this promoter to drive T7, which then activates a
30 very strong promoter (Studier et al., 1986). Such methods can be adapted for the TTSS of other bacteria.

[0155] Additionally, the number of needles in the bacterium can be increased by expressing the SPI-1 activator hilD. For example, by expressing HilD from the arabinose-inducible promoter PBAD (see Figure 17).

Ways to Optimize the N-terminal Secretion Signal

[0156] In some embodiments, the secretion system can be based on the 161 amino acid N-terminal secretion signal from SptP, which interacts with the SicP chaperone (Fu and Galan, 1998). This pair was chosen because it has been extensively studied and demonstrated to provide specificity towards the SPI-1 TTSS (Lee and Galan, 2004). However, it is somewhat disadvantageous as it is large and full-length SptP is only moderately secreted into cultures (Lee and Falkow, 2004). Thus additional secretion signals are contemplated.

[0157] For instance, two additional secretion signals can be cloned at the N-terminal end of ADF-3 and DH. First, one can clone the 15 amino acid leader from InvJ, which has been shown to be necessary and sufficient for secretion (Russman et al., 2002). InvJ is involved in needle assembly and is expressed into the media at high titer (Lee and Falkow, 2004). However, it is unclear whether InvJ secretion continues once the needle is constructed. A second secretion signal can be taken from the SopE effector, which interacts with the chaperone InvB (Lee and Galan, 2003). This effector is secreted in large amounts into the media. The first 100 amino acids can be used as the signal because it has been shown to include the chaperone binding domain (Lee and Galan, 2003). The SipA effector is secreted into the media at high titer, but its chaperone is SicA and, in some embodiments, might interfere with the SicA/InvF genetic circuit that we are using to induce.

[0158] In some embodiments, different signal-chaperone pairs may be optimal for secreting different types of proteins. For example, the SicP chaperone maintains proteins in a partially unfolded state, which may be important for the secretion of folded proteins (Stebbins and Galan, 2001). However, biopolymers do not have significant internal structure and may not require active maintenance in an unfolded state. In this case, a shorter signal, such as InvJ, may be adequate. The DH domain, which is a stable and folded protein, helps assess whether a secretion signal requires that a protein be unfolded.

[0159] We have explicitly measured different tags and tag-chaperone combinations. Of these, the SicP-SptP pair is particularly preferred. See figure 18)

Use of Directed Knockouts to Remove Competitive Secretion Systems, Effectors, and

Extra-cellular Proteases.

[0160] In some embodiments, the $\Delta invE$ strain is used. This strain significantly increases the titer of protein in the culture media. In other embodiments, additional directed knockouts
5 are contemplated to increase the titer of expressed protein. In such embodiments, genes can be knocked out to remove competitive secretion systems, effectors, and proteases.

[0161] For instance, the SPI-1 TTSS is co-regulated with the SPI-2 system and flagella (Deiwick et al., 1998; Eichelberg and Galan, 2000). There are regulatory interactions that ensure mutually exclusive interactions between systems. For example, when flagella are
10 expressed, SPI-1 is repressed. All three of these systems are expressed at different stages of growth in LB media. Cells that express one system do not express the other, which leads to a smaller fraction expressing SPI-1. To increase this fraction, the *flhD/C* master regulators of flagella assembly (Liu and Matsumura, 1994) and the *ssrAB* two-component system that controls SPI-2 (Feng et al., 2003) can be knocked out. Embodiments with these knockouts
15 have the added benefit of reducing the amount of proteins in the supernatant. Finally, in the current strain, SPI-1 effectors remain encoded in the genome. In embodiments having a knockout of the *sptP* and *sopE* effectors, these proteins will not compete for secretion. In additional embodiments, the *invJ* and *sipB/C* effectors are preferably left intact because of their involvement in the construction and function of the needle (Collazo and Galan, 1996;
20 Kubori et al., 2000).

Ways to Optimize the strength of the *sicA* promoter.

[0162] Embodiments employing the *sicA* promoter are contemplated as it is a relatively
25 strong promoter that is turned on late in the growth stage. However, other strong promoters can be integrated into the pertinent control circuits. For instance, synthetic genetic circuits can be constructed to increase expression from the *SicA* promoter. Instead of directly expressing the heterologous gene, the *sicA* promoter can be set to drive the expression of the T7 transcriptional activator. The heterologous protein then can be expressed from a very
30 strong T7 promoter (Studier et al., 1986). This circuit functions like an 'amplifier' to increase expression. Expression would still be linked to the completion of the TTSS structure and no chemical inducer would be required. Another advantage of using T7 is that it is not subject to rho-termination, making it less prone to producing protein truncation products (Fahnstock et al., 2000).

[0163] Overexpression can be balanced against other variables for further improvements. If the protein is particularly toxic or tends to particularly forms inclusion bodies, then higher levels of over expression could slow growth. In addition, with too high an overexpression, the secretion systems could be 'maxed out,' in which case the number of needles and secretion speed would be the rate determining step. Or, for instance, the chaperone *sicP* could be saturated and unable to deliver a higher concentration of effector to the needle. Maximizing yield may therefore require additional genetic manipulations to bolster the limiting factors.

Manufacture of Natural Biopolymers

- 10 [0164] DNA optimization and expression of natural biopolymers is generally contemplated as an important subgroup of proteins of interest. These proteins include the natural protein-based biopolymer genes of Table 1.

Table 1: Polymers to be synthesized

Name	Organism	Length
Dragline Silks		
ADF-3	<i>Araneus diadematus</i>	410
ADF-4	<i>Araneus diadematus</i>	410
NDF-3	<i>Nephila clavipes</i>	718
NDF-4		627
fibroin 1	<i>Euagrus chioseus</i>	734
fibroin 2	<i>Dolomedes tenebrosus</i>	691
dragline fibroin	<i>Euprosthops</i>	284
fibroin	<i>Antheraea pernyi</i>	724
fibroin 4	<i>Plectreuryx tristis</i>	1814
Ampullate Spidroins		
spidroin 1	<i>Nephila clavipes</i>	387
minor ampullate silk	<i>Nephila clavipes</i>	251
spidroin 1	<i>Kukulcania hibernalis</i>	760
spidroin 2-1	<i>Kukulcania hibernalis</i>	185
ADF-2	<i>Araneus diadematus</i>	294
major spidroin 1	<i>Argiope trifasciata</i>	648
major spidroin 1	<i>Latrodectus hesperus</i>	1065
ampullate spidroin	<i>Agelenopsis aperta</i>	847
Flagelliform Silks		
flagelliform silk	<i>Nephila clavipes</i>	907
ADF-1	<i>Araneus diadematus</i>	360
flagelliform silk	<i>Argiope trifasciata</i>	1002
Tubuliform Spidroins		

tubuliform spidroin	Nephila clavipes	592
tubuliform spidroin	Argiope auratia	376
tubuliform spidroin 1	Lactrodectus mactans	225
tubuliform spidroin 1	Deinopis spinosa	815
Non-spider Silks		
light chain	Bombyx mori	262
heavy chain	Bombyx mori	633
sericin 1A	Bombyx mori	779
heavy chain	Galleria mellonella	443
silk gum protein 1	Galleria mellonella	115
silk gum protein 2	Galleria mellonella	220
light chain	Galleria mellonella	267
Animal Polymers		
elastin	Homo sapiens	757
elastin microfibril 1	Homo sapiens	1016
fibulin	Homo sapiens	448
collagen	Homo sapiens	1678
abductin	Argopectin irradians	126

These polymers include dragline proteins from different spiders. Additional fibroins are included that represent threads with different material properties. For example, the web radial threads and safety line (ampullate), sticky spiral threads (flagelliform), and egg sack (cylindrical) silks all have different physical properties (Dicko et al., 2004). The genes from other species, such as moths and silk worms can be included. In particular, the extreme repetitive nature of silk worm silk has made it a hard target for recombinant technologies. The rubber-like resilin from *Drosophila* can also be expressed (Elvin et al., 2005). Other interesting biopolymers can be included from non-insect species. Mollusk abductin, which is an extremely strong glue-like protein with be synthesized (Bochicchio et al., 2005). Finally, interesting human biopolymers that make up the extracellular matrix – elastin, collagen, and elastin microfibril proteins – have applications in the coating of medical devices (Mongiat et al., 2000). Such proteins may be used as a fusion protein having the short secretion tag still attached or as a protein in which the tag has been removed.

[0165] The biopolymer genes can be codon optimized for eubacterial expression, while minimizing the DNA repetitiveness and mRNA secondary structures. Gene optimization and chemical synthesis are commercially available (e.g., GeneDesigner by DNA 2.0) to optimize sequences for translation. The sequence is then broken into oligonucleotides which are assembled either by annealing and ligation or annealing and extension (Prodromou and Pearl 1992; Yo et al. 2003).

[0166] The simplest way to design a DNA sequence from an amino acid sequence is to assign the most abundant codon to all instances of that amino acid in the sequence. Codon usage preference in a gene is often measured by Codon Adaptation Index (CAI score). The CAI score for such a construct is 1.0, *i.e.* in each case only the most abundant codon is used.

5 This 'one amino acid – one codon' or 'CAI=1.0' approach has several drawbacks. First, a strongly transcribed mRNA from such a gene can generate high codon concentrations for a subset of the tRNA populations, resulting in imbalanced tRNA pool, skewed codon usage pattern and increased translational error (Kurland and Gallant 1996). Heterologously expressed proteins may be produced at levels as high as 60% of total cell mass, making an
10 imbalance tRNA pool a significant problem resulting in reduced growth due to tRNA depletion (Gong M, Gong F et al. 2006) and increased frameshift due to translational pausing at the ribosomal A-site (Farabaugh and Björk 1999). Second, with no flexibility in codon selection, it is impossible to avoid repetitive elements and mRNA secondary structures in the gene. Severe repetitive elements can affect the genetic stability of a gene and may lead to
15 excision through recombination. Third, it is often desirable to incorporate or exclude sequence elements such as restriction sites from the sequence to facilitate subsequent manipulations. These modifications are impossible to accommodate if the codon usage is rigidly fixed.

[0167] In contrast to the 'CAI=1.0' method, Gene Designer optimizes genes for expression
20 by using a codon usage table in which each codon is given a probability score based on the frequency distribution of the codons in the genome normalized for every amino acid.

Candidate sequences are generated *in silico* using a Monte Carlo algorithm by selecting codons based on the probabilities obtained from the codon usage table, with codons below the threshold value (default is 10%) excluded from consideration. Each designed sequence is
25 then passed through subsequent iterations to ensure a match with additional design criteria such as filtering out mRNA secondary structures and DNA repeats, eliminating or incorporating restriction sites and avoiding methylation sites that overlap methylation sensitive restriction sites (Gustafsson, Govindarajan et al. 2004). The local context of a codon can influence the protein expression levels. Back in the early 1980s it was shown that the
30 efficiency of the UAG stop codon in *E. coli* is typically decreased in the presence of a 3' adenine and increased in the presence of a 3' cytidine (Bossi and Roth 1980; Miller and Albertini 1983). Gene Designer avoids known codon context issues by omitting the use of rare codons and filtering out runs of C's and G's.

[0168] Gene Designer does not utilize advanced RNA folding calculation software such as the popular mFold (Zuker 2003) as these types of software are designed to calculate RNA secondary structures for naked RNA. The translated mRNA within an ORF is in fact densely covered by ribosomes. Chemical footprinting of mRNA-ribosome complexes show that up to 20 codons (60 bases) are covered by a single translating ribosome (Green and Noller 1997), and the ribosomes are translating at ~18 codons (54bp)/sec with one ribosome initiating translation every ~2 second leaving only ~50 mRNA bases available between translating ribosomes for folding an mRNA secondary structure. During translation, a stem-loop structure in the coding part of the mRNA does not hinder the progress of the translational machinery, and actively translating ribosomes can break up such structures, either by the energy driven translation process itself or by the support of RNA helicases (Kurland et al. 1989; Iost and Dreyfus 1994).

[0169] Backtranslation with Gene Designer is performed in 2 stages. First a sequence encoding the desired amino acid sequence is selected by choosing each codon probabilistically using a codon bias table appropriate for the expression organism. Second an evolutionary algorithm is employed to remove excluded restriction sites, RNA secondary structure and repeated sequence elements. This algorithm compares the designed sequence with the additional constraints (such as the longest permitted repeat, the restriction enzyme recognition sequences that should not occur) and identifies codons that are part of regions that do not conform to the required specifications. These codons are then independently replaced by synonymous codons, again selected probabilistically from the codon bias table. A replacement that brings the sequence design closer to conforming to the additional constraints is accepted, otherwise it is rejected. This process is iterated hundreds or thousands of times until the constraints are met.

Assaying Optimized Genes for Expression and Secretion.

[0170] After each gene is synthesized, it can be placed in the *Salmonella* secretion strain. A FLAG-tag can be synthesized at the end of each fibroin, such that the concentration can be determined using a Western analysis. The secretion assays can be performed and the total protein can be quantified in the pellet lysate and supernatant. This can provide the total amount of protein produced and secreted, which can be used to determine if low yields are an expression or secretion problem.

Assaying for expression and secretion of each biopolymer or protein in *Salmonella* expression system or other expression systems.

[0171] Methods of analyzing proteins in fluid medium are well known to one of ordinary
5 in the art. They can be based upon antibody binding assays, competitive displacement assays, gel electrophoresis and staining assays, column chromatography, and the like. The proteins can be tagged with a fluorescent protein label such as GFP to facilitate their detection.

Preferably, the label can be joined to the protein by an amino acid sequence specifically
subject to hydrolysis by chemical means or by contact with a predetermined protease to
10 release the unlabeled protein.

Combining Natural and Synthetic Genetic Regulation to Create a Programmable Microbial Biopolymer or Heterologous Protein Factory.

[0172] In some embodiments, the SPI-1 initiation regulator hldD can be knocked out and
15 placed under inducible control by use of constructed hybrid SPI-1 promoters that are IPTG-inducible. Preferably, a protease site (e.g., a TEV protease site) is placed between the N-terminal secretion tag and the biopolymer or heterologous protein gene. A late stage TTSS promoter can be used to control TEV expression.

[0173] In one embodiment, the SPI-1 regulation can be used to program a series of events
20 into the bacterium, where a protein is first secreted and then modified. Both of these events may occur in the supernatant or growth medium. For example, a system can express a heterologous protein, and then automatically cleave the secretion signal. This can be done by placing the protein under the control of the *sicA* promoter, and then the protease under the control of a promoter that turns on at a higher cell density. The protease can be exposed to the
25 extracellular environment and cleave the N-terminal signal from the secreted protein. This system can provide a microbial biopolymer factory where each component is expressed in a precise, temporally regulated order.

[0174] In other embodiments, IPTG-inducible SPI-1 promoters are used. The SPI-1
regulatory cascade can be exploited to control the expression of different genes at different
30 stages of growth. Towards this goal, one can construct a set of IPTG-dependent SPI-1 promoters. The addition of IPTG to the media can unlock all of the promoters, but they retain their timing. In this way, the entire 'assembly line' can be activated at once and expression can be avoided during the growth and maintenance of the strain.

[0175] In some embodiments, inclusion of lacO binding sites up and downstream of an arbitrary promoter to repress its activity is contemplated (Law et al., 1993; Muller et al., 1996). A component of this repression is DNA looping. When the two binding sites are 70, 81, 92, 115, 150, or 206 nucleotides apart, this repression is optimal (Muller et al., 1996).

5 Synthetic promoters can be constructed by flanking SPI-1 and other *Salmonella* promoters by lacO binding sites and testing to determine if the promoter can be induced by IPTG. To obtain a range of on-times, IPTG-inducible *sicA*, *ssaG*, and *spvA* promoters can be constructed (Grob et al., 1997; McKelvie et al., 2004).

[0176] In other embodiments, automatic cleavage of an N-terminal polypeptide signal using a membrane-bound or other protease is contemplated. In some such embodiments, a TEV protease cleavage site can be placed between the N-terminal secretion signal and the heterologous protein. The TEV protease is small, well-characterized, and only leaves a single alanine attached to the protein. In further embodiments, the protease may be fused into the AIDA presentation system (Maurer et al., 1997), such that it remains tethered to the outer
15 membrane of the bacterium. This system has been shown previously to be functional in *Salmonella* (Rizos et al., 2003) and it is able to present large active enzymes to the extracellular environment (Latteman et al., 2000). The TEV-AIDA construct can be placed under the control of the *ssaG* promoter, which turns on in a subfraction of the cell population at a late stage of growth. The assay to monitor this system can be run a time course of
20 secreted protein and track the products by western blot. A SptP-DH protein, for instance, may be expressed and secreted and then the TEV protease automatically cleaves the 167 amino acid N-terminal signal. The IPTG-inducible promoters allow the entire system to be induced during growth such that the protease and heterologous protein do not accumulate during the growth and maintenance of the strain.

25 [0177] In some embodiments, cellulases or other enzymes are the heterologous protease to be secreted. In these embodiments, the bacteria can be engineered and the heterologous enzymes to be secreted selected according to a predetermined nutrient supply. For instance, the biomass converting cellulases, hemicellulases, and glycosyl hydrolases represent a naturally-secreted enzyme mixture that is of significant industrial interest. These enzymes
30 are secreted by fungi and bacteria in the course of breaking down plant biomass as a carbon source. These enzymes have been produced industrially to convert biomass to fermentable sugars to produce ethanol as an inexpensive biofuel. Cellulase production may be the most expensive step during ethanol production from cellulosic biomass, in that it can account for

approximately 40% of the total cost (Spano et al., 1975). Significant cost reduction is required in order to enhance the commercial viability of cellulase production technology.

[0178] A cellulosic enzyme system consists of three major components: endo- β -glucanase (EC 3.2.1.4), exo- β -glucanase (EC 3.2.1.91) and β -glucosidase (EC 3.2.1.21). The mode of action of each of these being:

- Endo-p-glucanase, 1,4- β -D-glucan glucanohydrolase, CMCase, Cx: "random" scission of cellulose chains yielding glucose and cello-oligo saccharides.
- Exo-P-glucanase, 1,4- β - D-glucan cellobiohydrolase, Avicelase, C1: exo-attack on the non-reducing end of cellulase with cellobiose as the primary structure.
- β -glucosidase, cellobiase: hydrolysis of cellobiose to glucose.

[0179] Production of these enzymes in a heterologous microbial system would be advantageous in three general areas. First, a controlled, high-titer system for producing secreted cellulase could bring down the cost of this industrial enzyme. Second, developing an easily transformable bacterial system for expressing of secreted cellulase could increase the through-put of a screen for more efficient enzymes. Finally, developing a genetically-tractable cellulase-secreting microbe is a significant step toward consolidated bioprocessing (CBP), a processing strategy for consolidating cellulosic biomass conversion into a single-step process. The CBP strategy could dramatically bring down the cost of valuable fermentation products.

[0180] These methods can provide an ultra low-cost process to produce an anti-malarial drug in microbes. Amyris Biotechnologies Inc. has engineered an *E. coli* strain that contains a nine-enzyme pathway much like the biosynthetic pathway in the plant *Artemisia annua*, responsible for the production of the anti-malarial drug, artemisinin (Martin et al., 2003). This engineered cell is capable of producing up to 20 g/L of the anti-malarial drug precursor in high-density fermentations. Shake flask experiments using the same system is *Salmonella enterica* have yielded similar results.

[0181] By outfitting such engineered cells with cellulase-secreting machinery, one can use plant waste as a carbon source for the production of low-cost antimalarial drugs and other drugs. As a proof of concept experiment, we propose to engineer *S. enterica* to secrete each of the cellulase components and/or a mixture. Demonstrating secretion, one can test the ability of these strain to co-metabolize cellulose and its constituents and use cellulose as a

sole or principal carbon source for growth. With the ability to use some form of cellulose as a carbon source, one can integrate the artemisinin and other precursor pathways to demonstrate the ability to produce drugs from cellulose, a low-cost agricultural waste product.

- 5 [0182] The following examples are intended to exemplify, but not to limit, various aspects of the inventive technology.

EXAMPLES

[0183] Natural silks are abundant biomaterials that span a remarkable diversity of physical properties (Vollrath, et al. 2001). The silk recovered from silk worms (*Bombyx mori*) is the most common agricultural source of material. However, many spiders and moths produce
10 silks with superior and desirable properties. For example, the dragline silk of spider webs is extremely strong yet remains highly elastic. In addition, humans also have silk-like proteins, which help form the structure of connective tissues. Together, these materials have a number of uses including medical device implants and high strength threads (Wang, Y., et al. 2006; Lewis, 2006). Unlike silk worms, the natural sources of these silks are not conducive to large-
15 scale agriculture, therefore requiring production in a recombinant host (Lewis, 2006).

[0184] The production of recombinant native silk proteins is complicated by several problems. First, the genes themselves are often unstable due to highly repetitive regions of DNA that results in frequent homologous recombination (Arcidiacono, S., et al. 1998).
20 Second, the codon usage in silk genes is not optimized for expression outside of the specialized cells in the silk gland (Rising et al., 2005; Prince et al. 1995). Rare codons result in ribosome pausing and early truncation producing a 'ladder' of incomplete protein products (Fahnestock et al., 1997). Third, the sequences are highly enriched in GC content, which can cause unusual mRNA secondary structures that reduce expression. Finally, if the proteins are
25 highly expressed in the confined cell volume they can produce fibrils. This was demonstrated elegantly by Huemmerich and co-workers, where intracellular fibrils were observed by light microscopy when the spider silk monomer ADF-4 was expressed (Huemmerich, 2004).

[0185] We have solved the first three problems by computationally designing optimized genes and constructing them using chemical DNA synthesis (Villalobos et al., 2006). An
30 algorithm was developed to re-assign the codon usage to match the recombinant host while maintaining the complete amino acid sequence. This algorithm simultaneously optimizes

codons for eubacterial expression, reduces mRNA secondary structure, and minimizes repeats in the DNA sequence.

[0186] Four silk genes from the orb weaving spider *Araneus diadematus* were optimized and constructed using DNA synthesis. These genes are expressed in different silk glands and vary in their amino acid content and material properties (Vollrath et al., 2001). ADF-1 is expressed in the minor ampullate gland and forms the radial spokes of a thread, which have high tensile strength, but are inelastic (Gosline et al., 1999; Guerette et al., 1996). ADF-2 is expressed in the cylindrical gland (egg sacks) and has a sequence that is similar to human elastins. ADF-3 and ADF-4 are expressed in the major ampullate gland and form the extremely tough and elastic dragline, which forms the frame of the web. The wild-type DNA sequences of these genes contain rare codons as well as large repetitive regions (Figure 14). In the case of ADF-4, there are two sets of DNA repeat units >100 bp with exact identity. Each gene was computationally optimized to eliminate rare codons, repetitive units, and reduce mRNA secondary structure. The optimized genes shared only 24-29% of the wild type codons. The optimizations were performed on the full available amino acid sequence, including non-repetitive N- and C-terminal domains, which have a role in solubility and the self-assembly of fibrils (Jin et al., 2003).

[0187] To solve the final problem, the bacterial type III secretion system (TTSS) was harnessed to export the silk monomers before they can form fibrils (Figure 11). The TTSS encoded on *Salmonella* Pathogenicity Island 1 (SPI-1) forms a needle-like structure that crosses both membranes and protrudes ~35 nm from the cell surface (Marlovits et al., 2006). When fully expressed, there are up to 100 needles per cell (Kurbori et al., 1998). In the natural context the needle functions as a molecular syringe to inject effector proteins into host cells that facilitate invasion and pathogenesis. The SPI-1 TTSS is advantageous because it has been well-characterized and the needles are highly expressed under standard laboratory conditions (Lundberg et al., 1999). In addition, the TTSS is unique because it translocates polypeptides through both the inner and outer membranes. This is in contrast to the sec and TaT pathways, which deliver proteins to the periplasm (Wickner et al., 2005; Georgiou et al., 2005). Type II secretion can export proteins from the periplasm through the outer membrane; however, the secretion signal is difficult to identify and appears to be distributed throughout the protein, making heterologous secretion difficult (Polschroder et al., 2005; Francetic et al., 2005).

[0188] Prior to constructing the synthetic control system, we characterized the dynamics of the natural SPI-1 regulation (Figures 12a). The network is organized such that environmental signals are received by the network via three transcription factors (HilC, HilD, and HilA), that initiate a four-tiered transcriptional cascade (Lucas et al., 2000). Transcriptional cascades have been shown to delay the activation of promoters (Rosenfeld et al., 2003). To quantify the activation of the SPI-1 network, promoter-green fluorescent protein transcriptional fusions were created. Parallel cultures were grown and shifted into SPI-1 inducing conditions (LB-miller, 0.3M NaCl). Upon induction, the transcriptional activators (*hilD*, *hilC*) are turned on first, then the structural genes (*prg*), and finally the effectors (*sicA*) (Figure 12a). This order of expression is analogous to the construction of the flagellum (Kalir et al., 2001). An important difference is that the SPI-1 circuitry boosts the expression of effectors after the construction of the needle, in contrast to the repression of flagellin that occurs after the filament is complete.

[0189] *Salmonella* uses a genetic circuit to avoid the expression of effector proteins until secretion needles have been constructed and are functional (Figure 12a) (Kalir et al., 2001). This circuit governs the activation of the *sicA* promoter, which controls the transcription of effector and chaperone genes. This promoter is turned on by the InvF transcription factor, which is only active when bound to the SicA chaperone. Prior to the completion of the TTSS, the SicA chaperone is sequestered by the SipB/C effectors (Tucker et al., 2000). Once the TTSS is functional, SipB/C secrete, thus freeing the chaperone to activate the transcription factor. This positive feedback loop amplifies the expression of effectors to match the capability of the cell to export protein.

[0190] The SicA gene circuit forms the core of our synthetic genetic system (pCASP) to secrete heterologous proteins. The *sicA* promoter has a low basal transcription rate and increases 200-fold in activity once the TTSS is functional (Figure 12c). The *sicA* promoter drives the expression of the heterologous protein, which is fused to an N-terminal secretion signal from the SptP effector protein (Lee et al., 2004). A tobacco etch virus (TEV) protease site is added after the signal sequence, such that the secretion signal can be cleaved after export. The SptP signal sequence interacts with the SicP chaperone, which directs the SptP tagged protein to the SPI-1 needle ((Akedo et al., 2005; Stebbins et al., 2001). The SicP chaperone is overexpressed with the secretion-tagged heterologous protein. The wild-type ribosome binding sites are preserved for both SicP and SptP to ensure that the native ratio of chaperone:effector is maintained.

[0191] Prior to working with silk proteins, heterologous protein secretion was tested using a 24kD human protein (DH domain), which expresses well in *Salmonella* (Figure 13). A secretion assay was performed to determine the amount of exported protein (Supplementary Information) (Collazo et al., 1996). After the SPI-1 secretion apparatus is assembled, there is a steady increase in the amount of DH protein secreted (Figure 13b). After 8 hours of growth in SPI-1 media, 60 mg/L of protein was detected. When DH lacking the secretion tag was expressed using an inducible system, the protein was detected in the cell lysate, but not after a secretion assay.

[0192] Each optimized gene was inserted into the pCASP system and tested for secretion (Figure 14). Secretion yields of 7, 18, 5, 1 mg/L/8-hours were obtained for ADF-1, 2, 3, and 4. As was observed with the DH protein, the removal of the N-terminal secretion tag eliminated silk secretion. After isolation, the N-terminal SptP tag can be removed by *in vitro* TEV proteolysis, leaving only two residues (serine, glycine) on the secreted protein.

[0193] The secretion efficiency was calculated by determining the amount expressed inside the bacteria to the amount that is secreted. In the case of ADF-3, 5.2 mg/L is detected in the lysate, yielding a secretion efficiency of 50% in eight hours (supplementary information). When the N-terminal tag is removed, slightly more is detected in the lysate (5.9 mg/L), but much less is present in the supernatant (200 µg/L) and can only be detected after concentration.

[0194] This work demonstrates how a system-level understanding of network dynamics can drive cellular engineering. Here, we have applied this approach to the construction of a synthetic control system that interacts with the natural regulation controlling type III secretion. This enables the expression and secretion of full-length spider silk monomers, which are notoriously difficult to express in recombinant hosts. This system is capable of producing silks spanning different material properties, making it possible to further explore – and modify – the amazing diversity of these materials. More generally, this work demonstrates the capability of the TTSS to export heterologous proteins of biotechnological interest. The TTSS has the unique capability to deliver proteins through both membranes of gram negative bacteria. Through the construction of synthetic control systems, this organelle can be harnessed for many applications in biotechnology.

Supplementary Information for the Example

1. Materials and Methods

1.A. Plasmids and Strains

[0195] *Salmonella typhimurium* SL1344 was used for all secretion experiments (gift of Stanley Falkow, Stanford). A flagella knockout was created by deleting the FlhCD transcriptional activators [Genbank ID: 1253445 and 1253446] (Δ FlhCD::KanR) using the method of Datsenko and Wanner in the *Salmonella* genome. The pCASP plasmid [Genbank ID: bankit870464] was constructed based on pPROTet.133 backbone (Cm^R, ColE1) (BD Clontech). The *sicA* promoter (165 bases upstream of the *sicA* start codon), *sicP* gene and first 160 amino acids of *sptP*, including the start codon [Genbank ID: 1254401, (3030898 - 3022551)] were obtained by PCR of *Salmonella typhimurium* SL1344 genomic DNA. A TEV protease cleavage sequence (GENLYFQSG), flanked by glycines for flexibility, was inserted by PCR primer between the SptP tag and the HindIII site. Open reading frames for DNA were inserted between the HindIII and XbaI restriction sites. A FLAG epitope tag (DYKDDDDK) was introduced non-directionally at the XbaI site for detection by western blot. The DH plasmid lacking the N-terminal secretion signal and SicP (DH no tag) was constructed using the pBAD30 backbone with the DH ORF inserted between the KpnI and XbaI sites. Manipulation of plasmids was done in *E. coli* strains XL1, MC1061, and OmniMAX (DNA 2.0). Media was supplemented with 25 ug/mL Kanamycin, 30 ug/mL Chloramphenicol or 100 ug/mL Ampicillin as needed.

1.B. Reporter Plasmid Construction

[0196] The reporter plasmids are constructed based on the pPROTet Cm^r system (ColE1 ori) available from Clontech/BD (Cat# 631203). The SPI-1 promoters were cloned from *Salmonella enterica Typhimurium* SL1344 genomic DNA, transcriptionally fused with green fluorescent protein (*gfpmut3*) using SOEing PCR, and ligated into pPRO using the XhoI/XbaI restriction sites. A large upstream region associated with each promoter was cloned: *hilD* (-312), *hilC* (-432), *prgH* (-317), *sicA* (-166).

1.C. Secretion Assay

[0197] The secretion assay was performed as described previously (Collazo et al., Infect Immune, 1996). The TTSS was induced in high-salt LB (LB Miller + 7g NaCl achieving a total NaCl concentration of 0.3M) and uninduced in LB-Lennox (L). Cells were plated on L-broth agar plates from frozen stock and grown overnight on L-broth agar. Single colonies were picked and grown 10 hours overnight in 5 ml liquid L broth. The overnights were

diluted 1:100 into 5mL fresh L-broth cultures and grown for 2 hours at 250 rpm. The cultures were then diluted 1:10 into 50mL of inducing media in a non-baffled 250mL glass flask. The cultures were grown at 37°C for 8 hours at 160 rpm. Supernatants were harvested by spinning cultures at 3500g for 30 minutes followed by vacuum filtration through 0.45um cellulose acetate filter unit (Corning Cat# 430314). At this point sample proteins were either precipitated in 10% trichloroacetic acid (TCA) for 1 hour and recovered or unconcentrated supernatant samples were collected. Samples were prepared in SDS sample buffer under reducing conditions and boiled for 3-5 minutes before PAGE analysis. Precipitated protein preparations were used only in the Coomassie gel and MalE western blot. All other studies used unconcentrated culture supernatant. Lysis is not detectible after the secretion assay, as determined by a Western blot of the periplasmic MalE protein and a Coomassie gel (Figure 24). For the positive lysis control a 50mL culture of saturated wild-type *Salmonella* were by pelleted at 3500xg for 30 minutes. The pellet was resuspended in 5mL of 25mM Tris 100mM NaCl buffer (pH 7.4), the resuspension was freeze-thaw cycled once at -80C. After thawing a catalytic amount of lysozyme was added and allowed to digest for 15 minutes. Finally the sample was sonicated for 2 min at 30% amplitude on ice before pelleting insoluble debris (5000xg 5 min).

1.D. Protease Cleavage

[0198] Samples for TEV proteolysis were prepared by collecting supernatants and combining undiluted sample with concentrated TEV sample buffer and recombinant TEV protease (Invitrogen Cat# 12575-015). TEV digests were run for 1 hour at room temperature before addition of SDS sample buffer under reducing conditions and analyzed by PAGE.

1.E. Detection

[0199] All heterologous proteins of interest were engineered to contain a C-terminal FLAG epitope tag (DYKDDDDK). Samples of supernatant were run on 10% or 12% polyacrylimide gels under reducing conditions and western blots performed using the FLAG antibody (Sigma Cat# F3165). For the MalE lysis study, an anti-MalE antibody (Genetex Cat# GTX20065) was used. A standard serial dilution of bacterial alkaline phosphatase with a c-terminal FLAG tag (Sigma Cat# P-7457) and of known concentration was run to allow quantification via standard curve.

1.F. Gene Synthesis

[0200] DNA 2.0 used in-house developed software (GeneDesigner) to optimize sequences for expression and translation. The degeneracy of the genetic code enables many alternative nucleotide sequences to encode the same protein. The frequencies with which different codons are used by different organisms and different types of genes vary significantly and are correlated to the concentration of the corresponding tRNA population in the cell. Each codon is given a probability score based on the frequency distribution of the codons in the genome normalized for every amino acid. Codons in the synthetic gene are then assigned from this table to create a new gene sequence. Gene Designer filters out (or flags, if it can not be avoided) any mRNA structure with double-stranded RNA stem of 12 bp or more. This feature is included to ensure that oligonucleotides used in the gene synthesis process do not predominantly self-anneal during gene assembly. Finally the sequence is broken into oligonucleotides, which are assembled either by annealing and ligation or annealing and extension.

1.G. Calculation of Sequence Entropy

[0201] The synthetic genes were optimized to reduce the presence of low abundance codons and DNA repetitiveness. To visualize the increase of DNA sequence diversity, the repeat units for each of the spider silks were manually aligned and a sequence entropy is calculated (Figure 25a). The sequence entropy S is defined as

$$S = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^4 p_i(j) \ln p_i(j)$$

where N is the length of the repeat unit and $p_i(j)$ is the probability that base j (A, T, G, C) occurs at position i . The maximum of this function (when all four bases are equally represented at each position) is ~ 0.6 .

1.H. *Salmonella* GFP-Reporter Growth Experiments:

[0202] All ODs were measured using a Cary 50 Bio spectrophotometer. Bacteria transformed with a reporter plasmid are grown overnight (~ 13 hours) in 5ml L broth (Difco, Lennox) supplemented with 34.4 $\mu\text{g/ml}$ Chloramphenicol. The overnight is then diluted 1:500 into 5ml fresh L-broth and grown 150 minutes at 37°C to $\text{OD}_{600} \sim 0.15$. A 2 ml aliquot is then added to 50 ml of inducing media (7g/L NaCl added to Difco Luria Bertani, Miller) in a non-baffled 250 ml flask. The cultures are grown in a shaker at 37°C at 160 rpm. During growth, 1 ml aliquots are taken every 20 min and the OD_{600} is measured, the cells are spun

down, resuspended in 200 μ l PBS with 2 mg/ml Kanamycin and put on ice to stop *gfp* expression. The data shown in Figure 23 includes data from four growth experiments performed on different days.

1.I. Flow Cytometry

5 **[0203]** Flow cytometry data was obtained on a BD FACSCalibur system as part of the UCSF core facility. Each bacterial dataset consists of at least 30,000 cells. The data is gated by forward and side scatter to observe bacteria-sized particles. Data analysis is performed using the WINMIDI program.

Secretion System Plasmid Map

10 **[0204]** The expression plasmid for heterologous secretion (pCASP) is shown Figure 15. A
 (Genbank:). The plasmid is based on the pPRO backbone with a ColE1 ori and Cm^R. The
 sicA promoter drives the expression of the chaperone SicP as well as the protein to be
 secreted. The secreted protein is fused to the N-terminal 160 amino acids of SptP. A TEV
 protease site is included so that the tag can be cleaved post-secretion. The plasmid is designed
 15 such that the HindIII/XbaI sites can be used to insert different genes to be secreted. The
 FLAG tag was synthesized as part of the ORF in the case of ADF-1, 2, 4. For ADF-3, it was
 added synthetically to the pCASP plasmid at the XbaI site.

3. Sequences of Synthetic Spider Silks

[0205] The amino acid and DNA sequences of the optimized synthetic spider silk genes are shown. In each amino acid sequence, the highly conserved repetitive unit is underlined. These repeat units are those used for sequence entropy statistics.

> ADF-1 amino acid sequence

HESSYAAAMAASTRNSDFIRNMSYOMGRLLSNAGAITESTASSAASSASSTVTESIRTYGPAAIFS

GAGAGAGVGVGGAGGYGOGYGAGA

GAGAGAGAGAGGAGGYGOGYGAGAAAAAGAGAGAAG

GYGGGSGAGAGGAGGYGOGYGAGSGAGAGAAAAAGASAGAAGGYGG

GAGVGAGAGAGAAAGGYGOSYGSAGAGAGAGA

GAAAAAGAGARAAGGYGGGYGAGA

GAGAGAAASAGASGGYGGGYGGGAGAGAVAGASAGSYGGAVNRLSSAGAASRVSSNVAAIASAGAAAL

PNVTSNTYSGVLSSGVSSSEALIOALLEVISALIHVLGSASIGNVSSVGVNSALNAVONAVGAYAG

```
> Synthetic ADF-1 gene (Native sequence, GI: 1263282)
```

CATGAATCTTCCTATGCTGCTGCAATGGCTGCTTCTACTCGTAATTCTGATTTTATCCGTAACATGAG

CTACCAGATGGGTCGTCTGCTGAGCAACGCCGGTGCCATTACCGAATCTACTGCAAGCAGCGCGGGCTT

CCAGCGCGTCCTCCACCGTTACCGAGTCTATTTCGCACGTATGGCCCGGCTGCGATCTTTTCTGGTGC

GGGCGCTGGCGCAGGCGTGGGTGTAGGTGGTGCCGGTGGTTACGGCCAGGGCTACGGTGCAG

GGGCGCTGGCCCAAGGCTGGGTGTAGGTGGTGGCGGTGGTTACGGCCAGGGCTACGGTGCAG
GCGCAGGTGCTGGTGGCGGGCGCCGGTGGCGGTGCTGGTGGCGCGGGGTGGCTACGGTCAGGGCTACGG

TGGCAGGTTGCCTTGGCGCGGGCGGGCTGGCGGCAGGCGGCAGGTGCAGCGGGCGGGTTACGGCGGGTGGGTTCTGG

TGCGGGGTGCTGCCGCGGCGGGCTGGCGCAGGGCGCAGGTGCAGCGGGCGGTTACGGCGGTGGTCTCGG
TGCTGGCGCTGGTGGTGGTGGTATGGCCAGGGTTACGGTGCAGGTTCTGGCGCGGGGTGCG

GGCGCTGCTGCGGCAGCTGGCGCATCCGCTGGTGCTGCTGGCGGCTATGGCGGTGGCGCAGGTGTTGG
 TGCAGGTGCGGGCGCGGGTGCAGGTGGTGGCTATGGCCAGAGCTATGGCAGCGGTGCTGGCGCAGGTG
 CGGGTGCTGGTGCGGCGGCTGCAGCTGGCGCTGGCGCACGTGCAGCGGGTGGCTACGGTGGTG
 GTTACGGCGCAGGCGCGGGCGCCGGTGTGGCGCCGCTGCTTCCGCTGGTGCCTCCGGTGGCTACG
 5 GTGGCGGTTACGGCGGTGGCGCGGGTGCAGGCGCCGTAGCTGGTGCCTCCGCGGGTCTTACGGCGGT
 GCGGTTAACCGTCTGTCTAGCGCAGGCGCGGCATCTCGTGTTCAGCAACGTGGCTGCCATCGCGTC
 TGCGGGTGCGGCTGCGCTGCCGAACGTAATCTCTAACATTTATTCTGGTGTGCTGTCCTCTGGTGTGT
 CTTCTTCTGAGGCGCTGATCCAGGCTCTGCTGGAAGTCATCTCTGCACTGATCCACGTGCTGGGTTCT
 GCCTCCATCGGTAACGTGTCTTCCGTTGGCGTTAACAGCGCACTGAATGCAGTGCAGAACGCCGTCGG
 10 CGCGTACGCTGGT

> ADF-2 amino acid sequence

GSQGAGGAGQGGYGAG
 GGGAAAAAAAAAVGAGGGGQGGLGSGGAGQGYGAGLGGQ
 15 GGASAAAAAAGGQGGQGGQGGYGGLGSQGAGGAGQLGYGAG
 QESAAAAAAGGAGGGGQGGLGAGGAGQGYGAAGLGGQGGAGQ
 GGGSGAAAAAGGQGGQGGYGGLG
 PQGAGGAGQGGYGGSLLQYGGQQAQAAAAASAAASRLSSPSAAARVSSAVSLVSNGGPTSPAALSSSI
 SNVVSQISASNPGLSGCDILVQALLEIISALVHILGSANIGPVNSSSAGQSASIVQSVYRALS
 20

> Synthetic ADF-2 gene (Native sequence, GI: 1263284)

GGTAGCCAAGGCGCAGGTGGTGCAGGTCAAGGTGGTTATGGTGCAGGCGGCGGTGGCGCTGCGGCAGC
 TGCTGCTGCAGCGGTAGGCGCGGGTGGCGGTGGTGCAGGCGGCGCTGGGTTCCGGCGGTGCGGGCCAGG
 GTTACGGCGCAGGCCTGGGCGGTCAAGGTGGCGCATCTGCGGCGGCTGCTGCGGCTGGTGGCCA
 25 GGGCGGTGAGGTGGCCAAGGTGGCTATGGCGGTCTGGGTTCTCAGGGCGCAGGCGGTGCTGGTC
 AGCTGGGCTATGGTGCAGGTGAGGAATCTGCAGCGGTGCCGCTGCCGCAGCGGGCGGCGCTGGTGGC
 GGTGGTCAGGGCGGCGCTGGGTGCGGGTGGCGCTGGCCAAGGTTACGGTGCCGCTGGCCTGGGCGGTCA
 GGGTGGTGCGGGCCAGGGCGGCGGCTCTGGCGCGGCGGCTGCGGCCGGTGGTCAAGGTGGTCA
 GGGCGGCTATGGTGGCCTGGGCCCCGCAAGGCGCGGGTGGTGCGGGCCAGGGTGGCTACGGTGGTGG
 30 TTCCCTGCAATACGGCGGTGAGGGTCAAGGCTCAGGCAGCTGCGGCATCCGCGGCGGCATCCCGCCTGT
 CTTCCCATCCGCAGCGGCACGTGTGTCTTCCGCTGTATCTCTGGTATCCAACGGCGGTCCGACCAGC
 CCGGCGGCACTGAGCTCTAGCATTTCCAACGTGGTATCTCAGATCTCTGCAAGCAACCCAGGCCTGTC
 TGGTTGCGATATCCTGGTTCAAGCCCTGCTGGAAATTATCTCTGCGCTGGTTCACATCCTGGGTTCTG
 CCAACATCGGCCCGGTTAACTCTAGCTCCGCCGGTCAGTCCGCATCCATTGTAGGTCAATCCGTATAC
 35 CGCGCTCTGTCT

> ADF-3 amino acid sequence

ARAGSGQGPGQGPGQGPGQGPGQGPYGP
 ASAAAAAGGYGPGSGQGPGSQQGPGQGPGQGPGQGPGQGPYGP
 40 ASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPGQGPYGP
 SSAAAAAGNGPGSGQGAGQGPGQGPGQGPG
 ASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPYGP
 ASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPGQGPYGP
 ASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPGQGPYGP
 45 ASAAAAAGGYGPGYQGPGQGPGQGPGQGPGQGPGQGPYGP
 ASAASAASGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPYGP
 ASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPGQGPGQGPYGP
 ASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPGQGPGQGPYGP
 ASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPGQGPYGP
 50 ASAAAGAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPGQGPGQGPGQGPYGP
 ASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQGPGQGPGQGPYGP
 PGGQGPYPGAASAASVGGYGPQSSSVVASAVASRLSSPAASSRVSSAVSSLVSSGPTKHAALSNT
 ISSVVSQVSASNPGLSGCDVLVQALLEVVSALVSILGSSSIGQINYGASAQYTMVQGQSVQAALA

> Synthetic ADF-3 gene (Native sequence, GI: 1263286)

GCCCCGCGGGGTCAGGCCAGCAGGGACCAGGTCAACAGGGCCCCGGGCCAACAAAGGCCCGGGTCAACA
 GGGTCCGTACGGTCCGGGTGCCAGCGCGGGCGGCCGCGGCCGAGGAGGGTATGGCCCTGGTAG
 CGGCCAACAGGGTCCGAGCCAGCAAGGCCCGGGCCAGCAAGGGCCGGGGGGCCAGGGGCCCTAC
 5 GGCCTTGGTGCCTCAGCTGCCGCAGCCGAGCTGGCGGTTATGGCCCGGGGTACAGGTACGAAGGGCC
 AGGCGGTCAAGGTCCCTACGGGGCCAGGCAGTAGTGCGGCAGCGGCTGCTGCCGGTGGTAACGGCC
 CGGGGTTCGGGCCAGCAAGGGGCGGGACAGCAGGGTCCAGGCCAACAAAGGCCCGGGTGCCTCCG
 CAGCGGCGGGCGGCCGCTGGTGGCTATGGCCCGGGTTCAGGCCAGCAGGGCCCGGGGCAGCAGGGCCCG
 GGTGGACAGGGTCCGTATGGCCCGGGGGCCAGTGCAGCGGCCGCGGCTGCTGGGGGCTATGGCC
 10 CTGGCTCAGGTACAGGGTCCGGGTCAACAAGGACCCGGCGGTCAAGGACCGTATGGCCCGGGTGC
 GTCCGCGGGCGGCTGCGGCGGCTGGAGGCTATGGTCCGGGAAGTGCCAACAGGGGCCCTGGACAGCAGG
 GTCCGGGTACAGCAGGGACCCGGTGGACAGGGCCCGTATGGGCCAGGCGCCTCTGCCGCAGCGGCGG
 CCGCAGGTGGGTATGGACCGGGGTACGGCCAGCAGGGTCTGGTACAGAGGGACCGGGCGGGC
 CAGGGCCCTTACGGCCCCGGCGCGTCAGCTGCAAGCGCTGCCTCGGGTGGCTACGGCCCGGGTTCG
 15 GTCAGCAGGGCCCGGGACAGCAGGGTCCGGGTACAGAGGACCGTATGGTCCGGGAGCTTCTGCTGC
 TGCCCGCGCGGGCGGGTGGTTATGGACCCGGCAGTGGCCAACAAGGTCCGGGGGCAGCAGGGTC
 CAGGTACAGAGGGCCAGGACAGCAGGGCCCTGGTGGCCAAGGACCGTACGGTCCCGGCGCAAGTGC
 GGCCGCTGCAGCTGCCGGAGGCTACGGTCCAGGTAGTGACAGCAAGGACCGGGTACAGAGGGCCCCG
 GTCAACAGGGGCCCGGGCCAGCAAGGCCCGGGCAGCAGGGACCTGGGCAGCAGGGTCCCGGGCAGCAA
 GGTCTTGGGCAACAGGGTCCGGGACAGCAAGGCCCTGGCGGCCAGGGTGCCTATGGGCCTGGTGCATC
 20 TGCCGCGGGCGGGCGCCGCGGGTGGGTACGGGCCGGGGAGCGGCCAGCAAGGTCCGGGCCAA
 CAGGGCCCCGGACAACAGGGTCTTGCCAGCAAGGACCTGGCCAGCAGGGGCCGGGACAACAAGGG
 CCCGGCCAACAAGGCCCAGGGCAACAAGGCCCGTACGGCCCTGGGGCCTCGGCAGCCGCGGCAGCGGC
 CGGCGGCTATGGCCCGGGCAGTGGTCAACAAGGTCCAGGCCAACAGGGCCCAGGGCAGCAGGGTCCGG
 GGGGTCAAGGTCCGTACGGACCGGGTGGCGCCTCGGCAGCGGTGAGTGTAGGCGGCTACGGACCTCAA
 25 AGCTCCTCTGTGCCAGTCGCCAGTGGGTGGCTAGCCGTCTGTCTAGCCCCGCCGCCAGCAGTCGTGT
 CAGCTCAGCCGTGTCTCTTAGTATCATCAGGACCGACTAAACACGCAGCCTTGTCAAACACCATTA
 GCAGCGTTGTCTCTCAGGTGTACGAGTAACCCGGGGCTGTGCGGTTGCGACGTCTGGTACAGGCC
 CTGCTGGAAGTGGTGAAGCGCCCTCGTGTCTATTCTGGGTTCTAGTTCATTGGCCAGATTAATACTATGG
 GCGAGTGCGCAATACACCCAAATGGTCCGACAATCTGTTGCGCAGGCACTGGCG

30 > ADF-4 amino acid sequence
A

GSSAAAAAASGSGGYGPENQGPSGPVAYGPGGPV
 SSAAAAAAGSGPGGYGPENQGPSGPGGYGPGGSG
 35 SSAAAAAASGPGGYGPSQGPSGPGGSGGYGPSQASGPGGPGAS
 AAAAAAASGPGGYGPSQGPSGPAYGPGGPG
 SSAAAAAASGPGGYGPSQGPSGPVYGPGGPG
 SSAAAAAAGSGPGGYGPENQGPSGPGGYGPGGSG
 SSAAAAAASGPGGYGPSQGPSGPGGSGGYGPSQGGSGPG
 40 ASAAAAAASGPGGYGPSQGPSGPYQGPSGPAYGPSASASVAASVYLRLQPRLEVSSAVSSL
 VSSGPTNGAAVSGALNSLVSQISASNPLSGCDALVQALLELVLSALVAILSSASIGQVNVSSVSQSTQ
 MISQALS

> Synthetic ADF-4 gene (Native sequence, GI: 1263288)
 GCAGGCTCTAGCGCCGAGCTGCCGCTGCTGCAAGCGGTAGCGGTGGTTACGGTCCAGAGAAC
 45 CAGGGTCCGTCCGGCCCAGTAGCATATGGCCCTGGTGGTCCAGTCTCTCCGCTGCTGCCGCAGC
 TGCTGCGGGCTCCGGTCCAGGTGGCTACGGTCCGGAAACAGGGCCCGTCTGGTCCGGGCGGTTATG
 GCCCGGGTGGCTCTGGTAGCTCTGCAGCGGCGGCAGCCGCGGCAGCGTCTGGCCCAGGTGGTTA
 CGGCCAGGCTCCCAGGGCCCGTCCGGTCCGGGCGGTAGCGGCGGTTATGGTCTGGTTCCCA
 50 GGGTGCAAGCGGCCCTGGTGGTCCGGGCGCATCTGCGGCAGCCGCCGAGCAGCGGCTGCGGCAAGCG
 GTCCGGGTGGCTACGGTCCGGGCAGCCAGGGTCCGTCTGGTCTTGGCGCCTACGGTCCAGGTGGCCCG
 GGTCTCTCCGCTGCAGCCGCGGCTGCGGCTGCGAGCGGTCTGGTGGCTACGGTCCGGGTAGC
 CAGGGTCTTCCGGTCCAGGTGTGTACGGCCCTGGTGGCCCGGGTCTCTGCTGCTGCAGCGG
 CGGCTGCTGGCTCTGGTCCGGGCGGTTATGGCCCGGAAACAGGGTCCGTCTGGCCCTGGTGGTTAC

GGCCCAGGTGGTTCCGGTTCCTCCGCTGCTGCGGCGGCAGCAGCTGCCAGCGGTCCAGGCGGTT
 ACGGTCCTGGCTCTCAAGGCCCGTCCGGCCCTGGCGGTTCCGGTGGCTATGGTCCGGGTCTCTCA
 GGGCGGTTCTGGTCCGGGCGCGAGCGCAGCTGCAGCGGCAGCCGCTGCATCTGGTCTGGCGGTTACG
 GTCCGGGTAGCCAGGGTCCATCCGGTCCGGGCTATCAGGGCCCGTCTGGCCCGGGTGCTTATGGTCCA
 5 TCCCCGAGCGCATCTGCGTCCGTGGCCGCTTCCGTCTATCTGCGTCTGCAACCGCGTCTGGAAGTTTC
 CTCTGCTGTTAGCAGCCTGGTTTCCAGCGGTCCGACTAACGGCGCTGCTGTCTCTGGCGCCCTGAACT
 CTCTGGTTTCCCAGATTTCTGCAAGCAACCCTGGTCTGTCTGGTTGCGACGCGCTGGTGCAGGCTCTG
 CTGGAAGTGGTTTCTGCGTGGTTGCAATCCTGAGCAGCGCAAGCATCGGTCAAGTTAACGTTTCTTC
 TGTCAGCCAGAGCACCCAAATGATTTCTCAGGCACTGAGC

4. Quantification of Silk Yields and Secretion Efficiency

[0206] Silk yields and secretion efficiency were measured by quantitative western blot. A standard secretion sample was collected for SptP tagged and tagless versions of ADF-3. The tagless ADF-3 was concentrated 20x by centrifugal filter (Amicon Cat# UFC801024) to
 15 allow visualization. The cell pellets from these experiments were washed with 10mL of PBS (pH 7.4) and pelleted for 30 minutes at 3500g and resuspended in 10mL PBS. The cells were lysed by addition of catalytic amounts of lysozyme (MP Biomedicals Cat#100834) and incubation for 15 minutes at room temperature followed by a 30 minute freeze/thaw cycle at -80C. Finally the lysate was sonicated for 2 minutes in 1 second pulses with 30% amplitude.
 20 The resulting mixture was spun to remove insoluble debris at 3500xg for 15 minutes. The soluble fraction was collected and frozen for quantitative western blot analysis. A standard ladder of bacterial alkaline phosphatase with a c-terminal FLAG tag (Sigma Cat# P-7457) of known concentration was run with each gel to allow quantification via a standard curve calculated using Photoshop (see Figure 16).

25 Literature Cited

Arcidiacono, S, Mello, C, Kaplan, D, Cheley, S, and Bayley, H (1998) Purification and characterization of recombinant spider silk in *Escherichia coli*, *Appl. Microbiol. Biotechnol.*, 49: 31-38.

30 Akeda, Y and Galan, JE (2005) Chaperone release and unfolding of substrates in type III secretion, *Nature*, 437: 911-915.

35 Bandiera, A, Taglienti, A, Micali, F, Pani, B, Tamaro, M, Crescenzi, V, and Manzini, G (2005) Expression and characterization of human-elastin-repeat-based temperature-responsive protein polymers for biotechnological purposes, *Biotechnol. Appl. Biochem.*, 42: 247-256.

40 Bini, E, Knight, DP, and Kaplan, DL (2004) Mapping domain structures in silks from insects and spiders related to protein assembly, *J. Mol. Biol.*, 335: 27-40.

Bochicchio, B, Jimenez-Oronoz, F, Pepe, A, Blanco, M, Sandberg, LB, and Tamburro, AM (2005) Synthesis of and structural studies on repeating sequences of abductin, *Macromol.*

Biosci., 5: 502-511.

Bossi, L. and J. R. Roth (1980). "The influence of codon context on genetic code translation." *Nature* 286(5769): 123-8.

5

Bouadloun, F., T. Srichaiyo, et al. (1986). "Influence of modification next to the anticodon in tRNA on codon context sensitivity of translational suppression and accuracy." *J Bacteriol.* 166(3): 1022-7.

10 Carrier, M. J. and R. H. Buckingham (1984). "An effect of codon context on the mistranslation of UGU codons in vitro." *J Mol Biol.* 175(1): 29-38.

Cheng, L. and E. Goldman (2001). "Absence of effect of varying Thr-Leu codon pairs on protein synthesis in a T7 system." *Biochemistry* 40(20): 6102-6.

15

Chilcott, GS, and Hughes, KT (2000) Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar *typhimurium* and *Escherichia coli*, *Microb. Mol. Biol. Rev.*, 64: 694.

20 Cirillo, DM, Valdivia, RH, Monack, DM, and Falkow, S (1998) Macrophage-dependent induction of the *Salmonella* pathogenicity island 2 type III secretion system and its role in intracellular survival, *Mol. Microb.*, 30: 175-188.

25 Collazo, CM, and Galan, JE (1996) Requirement for exported proteins in secretion through the invasion-associated type III system of *Salmonella typhimurium*, *Infect. Immun.*, 64: 3524-3531.

30 Dal Pra, I, Freddi, G, Minic, J, Chiarini, A, and Armato, U (2005) De novo engineering of reticular connective tissues in vivo by silk fibroin nonwoven materials, *Biomaterials*, 26: 1987-1999.

Darwin, K.H., and Miller, V.L. (2001) Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in *Salmonella typhimurium*, *EMBO J.*, 20: 1850-1862.

35

Deiwick, J., Nikolaus, T., Shea, J.E., Gleeson, C., Holden, D.W., and Hensel, M. (1998) Mutations in *Salmonella* pathogenicity island 2 (SPI2) genes affecting transcription of SPI1 genes and resistance to antimicrobial agents, *J. Bacteriol.*, 180: 4775-4780.

40 Dicko, C, Knight, D, Kenney, JM, and Vollrath, F (2004) Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration effects, *Biomacromolecules*, 5: 2105-2115.

45 Eichelberg, K., and Galan, J.E. (2000) The flagellar sigma factor FliA (s28) regulates the expression of *Salmonella* genes associated with the Centisome 63 Type III secretion system, *Infect. Immun.*, 68: 2735-2743.

Elvin, CM, et al (2005) Synthesis and properties of crosslinked recombinant pro-resilin, *Nature*, 437: 999-1002.

50

- Fahnestock, SR, and Bedzyk, LA (1997a) Production of synthetic spider dragline silk protein in *Pichia pastoris*, *Appl. Microbiol. Biotechnol.*, 47: 33-39.
- 5 Fahnestock, SR, and Irwin, SL (1997b) Synthetic spider dragline silk proteins and their production in *Escherichia coli*, *Appl. Microbiol. Biotechnol.*, 47: 23-32.
- Fahnestock, SR, Yao, Z, and Bedzyk, LA (2000) Microbial production of spider silk proteins, *Rev. Mol. Biotechnology*, 74: 105-119.
- 10 Farabaugh, P. J. and G. R. Björk (1999). "How translational accuracy influences reading frame maintenance." *Embo J* 18(6): 1427-34.
- Feng, X., Oropeza, R., and Kenney, L.J. (2003) Dual regulation by phosphor-OmpR of *ssrA/B* gene expression in *Salmonella* pathogenicity island 2, *Mol. Microb.*, 48: 1131-1143.
- 15 Francetic, O. and A.P. Pugsley, *Towards the identification of type II secretion signals in a nonacylated variant of pullulanase from Klebsiella oxytoca*. *J Bacteriol*, 2005. **187**(20): p. 7045-55.
- 20 Fu, Y, and Galan, JE (1998) Identification of a specific chaperone for SptP, a substrate of the centisome 63 type III secretion system of *Salmonella typhimurium*, *J. Bacteriol.*, 180: 3393-3399.
- 25 Galan, J.E., and Collmer, A. (1999) Type III secretion machines: bacterial devices for protein delivery into host cells, *Science*, 284: 1322-1328.
- Galan, J.E., and Zhou, D. (2000) Striking a balance: Modulation of the actin cytoskeleton by *Salmonella*, *Proc. Natl. Acad. Sci. USA*, 97: 8754-8761.
- 30 Gao, X., P. Yo, et al. (2003). "Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences." *Nucleic Acids Res* 31(22): e143.
- 35 Georgiou, G. and L. Segatori, *Preparative expression of secreted proteins in bacteria: status report and future prospects*. *Curr Opin Biotechnol*, 2005. **16**(5): p. 538-45.
- Gatesy, J, Hayashi, C, Motriuk, D, Woods, J, and Lewis, R (2001) Extreme diversity, conservation, and convergence of spider silk fibroin sequence, *Science*, 291: 2603-2605.
- 40 Gong M, Gong F, et al. (2006). "Overexpression of *tnaC* of *Escherichia coli* Inhibits Growth by Depleting tRNA^{Pro} Availability." *J Bacteriol.* 188(5): 1892-8.
- Gonzalez de Valdivia, E. and L. A. Isaksson (2005). "Abortive translation caused by peptidyl-tRNA drop-off at NGG codons in the early coding region of mRNA." *FEBS J.* 272(20): 5306-16.
- 45 Gouy, M. (1987). "Codon contexts in enterobacterial and coliphage genes." *Mol Biol Evol.* 4(4): 426-44.
- 50 Gosline, JM, Guerette, PA, Ortlepp, CS, and Savage, KN (1999) The mechanical design of

spider silks: from fibroin sequence to mechanical function, J. Exp. Biol., 202: 3295-3303.

Green, R. and H. Noller (1997). "Ribosomes and translation." Annu Rev Biochem. 66: 679-716.

5

Grob, P, Kahn, D, and Guiney, D (1997) Mutational characterization of promoter regions recognized by the *Salmonella* Dublin virulence plasmid regulatory protein SpvR, J. Bacteriol., 179: 5398-5406.

10 Guerette, P.A., et al., *Silk properties determined by gland-specific expression of a spider fibroin gene family*. Science, 1996. **272**(5258): p. 112-5.

Gustafsson, C., S. Govindarajan, et al. (2004). "Codon bias and heterologous protein expression." Trends Biotechnol 22(7): 346-53.

15

Gutman, G. A. and G. W. Hatfield (1989). "Nonrandom utilization of codon pairs in *Escherichia coli*." Proc Natl Acad Sci U S A. 86(10): 3699-703.

20 Haraga, A., and Miller, S.I. (2003) A *Salmonella enterica* serovar *typhimurium* translocated leucine-rich repeat effector protein inhibits NF-kB-dependent gene expression, Infect. Immun., 71: 4052-4058.

Hagervall, T. and G. Bjork (1984). "Undermodification in the first position of the anticodon of supG-tRNA reduces translational efficiency." Mol Gen Genet. 196(2): 194-200.

25

Ham, JH, Bauer, DW, Fouts, DE, and Collmer, A (1998) A cloned *Erwinia chrysanthemi* Hrp (type III protein secretion) system functions in *Escherichia coli* to deliver *Pseudomonas syringae* Avr signals to plant cells and to secrete Avr proteins in culture, Proc. Natl. Acad. Sci. USA, 95: 10206-10211.

30

Ham, TS, Lee, SK, Keasling, JD, and Arkin, AP (2006) A tightly regulated inducible expression system utilizing the fim inversion recombination switch, Biotech. Bioeng, DOI 10.1002/bit. 20916.

35 Hayashi, CY, and Lewis, RV (2001) Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution, BioEssays, 23: 750-756.

Hayes, C., B. Bose, et al. (2002). "Stop codons preceded by rare arginine codons are efficient determinants of SsrA tagging in *Escherichia coli*." Proc Natl Acad Sci U S A 99(6): 3440-5.

40

Henaut, A. and A. Danchin (1996). Analysis and predictions from *Escherichia coli* sequences. *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. C. Neidhardt F, R. I. Curtiss, J. Ingraham et al. Washington, D.C, ASM press. 2: 2047-2066.

45 Hernandez, L.D., Pypaert, M., Flavell, R.A., and Galan, J.E. (2003) A *Salmonella* protein causes macrophage cell death by inducing autophagy, J. Cell. Biol., 163: 1123-1131.

Hinman, MB, Jones, JA, and Lewis, RV (2000) Synthetic spider silk: a modular fiber, TIBTech, 18: 374-379.

50

- Huemmerich, D, et al (2004a) Primary structure elements of spider dragline silks and their contribution to protein solubility, *Biochemistry*, 43: 13604-13612.
- 5 Huemmerich, D, Scheibel, T, Vollrath, F, Cohen, S, Gat, U, and Ittah, S (2004b) Novel assembly properties of recombinant spider dragline silk proteins, *Curr. Biol.*, 14: 2070-2074.
- 10 Ikemura, T. (1981). "Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. Coli* translational system." *J Mol Biol* 151(3): 389-409.
- 15 Ingraham, J. L., Maaloe, et al. (1983). Growth rate as a variable. Growth of the bacterial cell. Sunderland, MA., Sinauer Associates Inc.: 267-315.
- 20 Iost, I. and M. Dreyfus (1994). "mRNAs can be stabilized by DEAD-box proteins." *Nature* 372: 193-196.
- 25 Irwin, B., J. D. Heck, et al. (1995). "Codon pair utilization biases influence translational elongation step times." *J Biol Chem.* 270(39): 22801-6.
- 30 Jin, H., Q. Zhao, et al. (2006). "Influences on gene expression in vivo by a Shine–Dalgarno sequence." *Mol. Microbiol.* In Press.
- 35 Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M.G., and Alon, U. (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria, *Science*, 292: 2080-2083.
- 40 Kane, J., B. Violand, et al. (1992). "Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*." *Nucleic Acids Res* 20(24): 6707-12.
- 45 Kodumal, S. J., K. G. Patel, et al. (2004). "Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster." *Proc Natl Acad Sci U S A* 101(44): 15573-8.
- 50 Kubori, T, and Galan, JE (2002) *Salmonella* type III secretion-associated protein InvE controls translocation of the effector proteins into host cells, *J. Bacteriol.*, 184: 4699-4708.
- Kubori, T, Sukhan, A, Aizawa, S, and Galan, JE (2000) Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system, *Proc. Natl. Acad. Sci. USA*, 97: 10225-10230.
- Kubori, T., et al., *Supramolecular structure of the Salmonella typhimurium type III protein secretion system*. *Science*, 1998. **280**(5363): p. 602-5.
- 45 Kumar, D., C. Gustafsson, et al. (2006). "Validation of RNAi Silencing Specificity Using Synthetic Genes: Salicylic Acid-binding Protein 2 Is Required For Plant Innate Immunity." *Plant J.* 45(5): 863-8.
- 50 Kurland, C. and J. Gallant (1996). "Errors of heterologous protein expression." *Curr Opin Biotechnol* 7(5): 489-93.

- Lattemann, CT, Mauer, J, Gerland, E, and Meyer, TF (2000) Autodisplay: Functional display of active β -lactamase on the surface of *Escherichia coli* by the AIDA-I autotransporter, *J. Bacteriol.*, 182: 3726-3733.
- 5 Laub, M.T., McAdams, T.H., Feldblyum, T., Fraser, C.M., and Shapiro, L. (2000) Global analysis of the genetic network controlling a bacterial cell cycle, *Science*, 290: 2144-2148.
- 10 Laursen, B. S., H. P. Sørensen, et al. (2005). "Initiation of Protein Synthesis in Bacteria." *Microbiol. Mol. Biol. Rev.* 69(1): 101-123.
- Law, SM, Bellomy, GR, Schlax, PJ, and Record, MT (1993) In vivo thermodynamic analysis of repression with and without looping in lac constructs, *J. Mol. Biol.*, 230: 161-173.
- 15 Lazeris, A, et al. (2002) Spider silk fibers spun from soluble recombinant silk produced in mammalian cells, *Science*, 295: 472-476.
- 20 Lee, C.A., and Falkow, S. (1992) Identification of a *Salmonella-typhimurium* invasion locus by selection for hyperinvasive mutants, *Proc. Natl. Acad. Sci. USA*, 89: 1847-1851.
- Lee, SH, and Galan, JE (2003) InvB is a type III secretion-associated chaperone of the *Salmonella enterica* effector protein SopE, *J. Bacteriol.*, 185: 7279-7284.
- 25 Lee, SH, and Galan, JE (2004) *Salmonella* type III secretion-associated chaperones confer secretion-pathway specificity, *Mol. Microb.*, 51: 483-495.
- 30 Lewis, RV, Hinman, M, Kothankota, and Fournier, MJ (1996) Expression and purification of a spider silk protein: a new strategy for producing repetitive protein, *Protein Expression and Purification*, 7: 400-406.
- Lewis, R.V., *Spider silk: ancient ideas for new biomaterials*. *Chem Rev*, 2006. **106**(9): p. 3762-74.
- 35 Lilic M. et al. (2006) A Common Structural Motif in the Binding of Virulence Factors to Bacterial Secretion Chaperones. *Molecular Cell*. 21: 653-664.
- Lithwick, G. and H. Margalit (2003). "Hierarchy of sequence-dependent features associated with prokaryotic translation." *Genome Res* 13(12): 2665-73.
- 40 Liu, XY, and Matsumura, P (1994) The FlhD FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons, *J. Bacteriology*, 176: 7345-7351.
- 45 Looman, A. C., J. Bodlaender, et al. (1987). "Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in *Escherichia coli*." *EMBO J.* 6(8): 2489-92.
- Lucas, R.L. and Lee, C. A. (2001) Roles of hilC and hilD in regulation of hilA expression in *Salmonella enterica* serovar *typhimurium*, *J. Bacteriol.*, 183: 2733-2745.
- 50 Lucas, R.L. and C.A. Lee, *Unravelling the mysteries of virulence gene regulation in*

Salmonella typhimurium. Mol Microbiol, 2000. **36**(5): p. 1024-33.

Lundberg, U., et al., *Growth phase-regulated induction of Salmonella-induced macrophage apoptosis correlates with transient expression of SPI-1 genes*. J Bacteriol, 1999. **181**(11): p. 3433-7.

Majander K. et al. (2005) Extracellular secretion of polypeptides using a modified *Escherichia coli* flagellar secretion apparatus. Nature Biotechnology. 23: 475-481.

Marlovits, T.C., et al., *Structural insights into the assembly of the type III secretion needle complex*. Science, 2004. **306**(5698): p. 1040-2.

Marlovits, T.C., et al., *Assembly of the inner rod determines needle length in the type III secretion injectisome*. Nature, 2006. **441**(7093): p. 637-40.

Martin et al. *Nat Biotech* 21, 796 - 802 (2003)

Martin, SL, Vrhovski, B, and Weiss, AS (1995) Total synthesis and expression in *Escherichia coli* of a gene encoding human tropoelastin, Gene, 154: 159-166.

Maskarinec, SA, and Tirrell, DA (2005) Protein engineering approaches to biomaterials design, Curr. Opin. Biotech., 16: 422-426.

Mauer, J, Jose, J, and Meyer, TF (1997) Autodisplay: One-component system for efficient surface display and release of soluble recombinant proteins from *Escherichia coli*, J. Bacteriol., 179: 794-804.

McKelvie, ND, et al., (2004) Expression of heterologous antigens in *Salmonella typhimurium* vaccine vectors using the in vivo-inducible SPI-2 promoter ssaG, Vaccine, 22: 3243-3255.

McMillan, RA, Lee, TAT, and Conticello, VP (1999) Rapid assembly of synthetic genes encoding protein polymers, Macromolecules, 32: 3643-3648.

McNulty, D., B. Claffee, et al. (2003). "Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*." Protein Expr Purif 27(2): 365-74.

Mello, CM, Soares, JW, Arcidiacono, S, and Butler, MM (2004) Acid extraction and purification of recombinant spider silk proteins, Biomacromolecules, 5: 1849-1852.

Miller, J. H. and A. M. Albertini (1983). "Effects of surrounding sequence on the suppression of nonsense codons." J Mol Biol. 164(1): 59-71.

Mongiat et al., (2000) Self-assembly and supramolecular organization of EMILIN, J. Biol. Chem., 275: 25471-25480.

Moura, G., M. Pinheiro, et al. (2005). "Comparative context analysis of codon pairs on an ORFeome scale." Genome Biol. 6(3): R28.

Muller, J, Oehler, S., Muller-Hill, B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator, J. Mol. Biol., 257: 21-20.

- Murgola, E., F. T. Pagel, et al. (1984). "Codon context effects in missense suppression." J Mol Biol. 175(1): 19-27.
- 5 Nakamura, Y., T. Gojobori, et al. (2000). "Codon usage tabulated from international DNA sequences databases: status for the year 2000." Nucleic Acids Res. 28: 292.
- Pohlschröder M. et al. *Diversity and Evolution of Protein Translocation*. Annu. Rev. Microbiol. 2005, 59, p91-111.
- 10 Prince, JT, McGrath, KP, DiGirolamo, CM, and Kaplan, DL (1995) Construction, cloning and expression of synthetic genes encoding spider dragline silk, Biochemistry, 34: 10879-10885.
- 15 Prodromou, C. and L. H. Pearl (1992). "Recursive PCR: a novel technique for total gene synthesis." Protein Eng 5(8): 827-9.
- Reese, E.T. et al., J. Bacteriol., 59, 485-497 (1950)
- 20 Rising, A., et al., *Spider silk proteins--mechanical property and gene sequence*. Zoolog Sci, 2005. **22**(3): p. 273-81.
- Rizos, K, Latteman, CT, Bumann, D, Meyer, TF, and Aebischer, T (2003) Autodisplay: efficacious surface exposure of antigenic UreA fragments from Helicobacter pylori in
- 25 *Salmonella* vaccine strains, Infect. Immun., 71: 6320-6328.
- Rosenfeld, N. and U. Alon, *Response delays and the structure of transcription networks*. J Mol Biol, 2003. **329**(4): p. 645-54.
- 30 Rossman, KL, Worthylake, DK, Snyder, JT, Siderovski, DP, Campbell, SL, Sondek, J (2002) A crystallographic view of interactions between Dbs and Cdc42: PH domain-assisted guanine nucleotide exchange, EMBO J., 21: 1315-1326.
- Russmann, H, Shams, H, Poblete, F, Fu, XY, Galan, JE, and Donis, RO (1998) Delivery of epitopes by the *Salmonella* type III secretion system for vaccine development, Science, 281: 565-568.
- 35 Russmann, H, Kubori, T, Sauer, J, and Galan, JE (2002) Molecular and functional analysis of the type III secretion signal of the *Salmonella* enterica InvJ protein, Mol. Microb., 46: 769-779.
- 40 Sheller, J, Guhrs, K-H., Grosse, F, and Conrad, U (2001) Production of spider silk protein in tobacco and potato, Nature Biotechnology, 19: 573-577.
- 45 Shotland, Y., Kramer, H., and Groisman, E.A. (2003) The *Salmonella* SpiC protein targets the mammalian Hook3 protein function to alter cellular trafficking, Mol. Microb., 49: 1565-1576.
- 50 Shea, J.E., Hensel, M., Gleeson, C., and Holden, D.W. (1996) Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*, Proc. Natl.

Acad. Sci. USA, **93**: 2593-2597.

Shpaer, E. G. (1986). "Constraints on codon context in *Escherichia coli* genes. Their possible role in modulating the efficiency of translation." *J Mol Biol.* 188(4): 555-64.

Sorensen, M. A., C. G. Kurland, et al. (1989). "Codon usage determines translation rate in *Escherichia coli*." *J Mol Biol.* 207(2): 365-77.

Spano, L. et al., In "Proceeding of 2nd Annual Symp. on Fuels from Biomass" Ed. Shuster, W.W., 671-684 (1978) John Willey and Sons, New York

Sprengart, M. L., E. Fuchs, et al. (1996). "The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*." *EMBO J.* 15(3): 665-74.

Stebbins, CE, and Galan, JE (2001) Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion, *Nature*, 414: 77-81.

Stebbins, CE, and Galan, JE (2003) Priming virulence factors for delivery into the host, *Nature Reviews*, 4: 738-743.

Stenstrom, C. M., E. Holmgren, et al. (2001). "Cooperative effects by the initiation codon and its flanking regions on translation initiation." *Gene* 273(2): 259-65.

Studier, FW, and Moffatt, BA: Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 1986, 189: 113-130.

Sukhan, A., Kubori, T., Wilson, J., and Galan, J.E. (2001) Genetic analysis of assembly of the *Salmonella enterica* serovar *typhimurium* type III secretion-associated needle complex, *J. Bacteriol.*, 183: 1159-1167.

Takyar, S., R. P. Hickerson, et al. (2005). "mRNA helicase activity of the ribosome." *Cell* 120(1): 49-58.

Tian, J., H. Gong, et al. (2004). "Accurate multiplex gene synthesis from programmable DNA microchips." *Nature* 432(7020): 1050-4.

Tsubouchi, K, Igarashi, Y, Takasu, Y, and Yamada, H (2005) Sericin enhances attachment of cultured human skin fibroblasts, *Biosci. Biotechnol. Biochem.*, 69: 403-405.

Tucker, S.C. and J.E. Galan, *Complex function for SicA, a Salmonella enterica serovar typhimurium type III secretion-associated chaperone.* *J Bacteriol*, 2000. **182**(8): p. 2262-8.

Van Beek, JD, Hess, S, Vollrath, F, and Meier, BH (2002) The molecular structure of spider dragline silk: Folding and orientation of the protein backbone, *Proc. Natl. Acad. Sci. USA*, 99: 10266-10271.

Villalobos, A., et al., *Gene Designer: a synthetic biology tool for constructing artificial DNA segments.* *BMC Bioinformatics*, 2006. **7**: p. 285.

Vollrath, F, and Knight, DP (2001) Liquid crystalline spinning of spider silk, *Nature*, 410:

541-548.

Wang, Y., et al., *Cartilage tissue engineering with silk scaffolds and human articular chondrocytes*. Biomaterials, 2006. **27**(25): p. 4434-42.

5

Wickner, W. and R. Schekman, *Protein translocation across biological membranes*. Science, 2005. **310**(5753): p. 1452-6.

10

Yang, J, et al (2005) High yield recombinant silk-like protein production in transgenic plants through protein targeting, *Transgenic Research*, 14: 313-324.

Young, L. and Q. Dong (2004). "Two-step total gene synthesis method." *Nucleic Acids Res* 32(7): e59.

15

Zaslaver, A., Mayo, A.E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M.G., and Alon, U. (2004) Just-in-time transcription program in metabolic pathways, *Nature Genetics*, 36: 486-491.

20

Zierler, MK, and Galan, JE (1995) Contact with cultured epithelial cells stimulates secretion of *Salmonella typhimurium* invasion protein InvJ, *Infect. Immun.*, 63: 4024-4028.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." *Nucleic Acids Res* 31(13): 3406-15

25

U.S. Patent Application Publication No. 20060068469, March 2006.

[0207] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof can be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes. No reference to a publication herein should be construed as an admission that such is prior art.

WHAT IS CLAIMED IS:

1. A method of making a protein, said method comprising expressing the protein in a gram negative bacterium having a type III secretion system wherein the protein is heterologous to the bacterium and the protein is fused to a polypeptide that directs the protein to the secretion system, and wherein the protein is secreted via the secretion system into the medium of the bacterium.
2. The method of claim 1, wherein the bacterium is selected from the group consisting of *Salmonella*, *E. Coli*, *Yersinia*, *Shigella*, and *Bordetella*.
3. The method of claim 1, wherein the bacteria is *Salmonella typhimurium*.
4. The method of claim 1, wherein the bacterium is *Salmonella typhimurium* 1344
5. The method of claim 1, wherein the bacterium has a reduced ability to express an effector protein secreted by the secretion system.
6. The method of claim 5, wherein multiple genes encoding effector proteins involved in virulence have been deleted or silenced.
7. The method of claim 1, wherein the Type III secretion system is an SPI-1 secretion system.
8. The method of claim 1, wherein the bacterium has contact dependence genes, and said contact dependence genes have been knocked out to remove the contact dependence (requirement for a host cell) to increase secretion into the media.
9. The method of claim 1, wherein the heterologous protein is a therapeutic protein.
10. The method of any one of claims 1 to 9, wherein the heterologous protein is a human protein.
11. The method of claim 10, wherein the heterologous protein is isolated from the medium and formulated with a pharmaceutically acceptable carrier.

12. The method of claim 1, wherein the heterologous protein is a spider silk protein.
13. The method of claim 12, wherein the protein is an ADF protein.
14. The method of claim 1, wherein the heterologous protein is a silk worm silk protein.
15. The method of claim 1, wherein the protein is elastin.
16. The method of claim 1, wherein the protein is a fibroin or a protein-based biopolymer or an amino acid based macrobiopolymer.
17. The method of claim 1, wherein a tag is coupled to the protein by a sequence of amino acids which is cleavable by an enzyme and the secreted protein is contacted with the enzyme and cleaved by the enzyme.
18. The method of claim 17, wherein the enzyme is TEV protease and the sequence comprises the TEV recognition sequence.
19. The method of claim 1, wherein the protein is encoded by a gene wherein the DNA was structurally stabilized and optimized for expression in bacteria through codon optimization, mRNA minimization, and reduction of recombination frequency.
20. The method of claim 1, wherein the heterologous protein is expressed by a gene on a plasmid.
21. The method of claim 1, wherein the expression of the protein is controlled by a genetic circuit that links the activation of expression to the completion of the TTSS structure.
22. The method of claim 21, wherein the expression of the protein is under the control of a promoter
23. The method of claim 21, wherein the expression of the protein is under the control of the *sicA* promoter.
24. The method of claim 21, wherein the expression of the protein is under the control of a constitutive or inducible promoter.

25. The method of claim 1, where the expression of the protein is controlled by a SPI-1 promoter.
26. The method of claim 25 wherein the promoter is selected from the group consisting of hilC, hilD, hilA, invF, sopE, and prgH.
27. The method of claim 1, wherein the bacterium is *Salmonella* and the tag is a tag of a *Salmonella* effector.
28. The method of claim 27, wherein the tag is an SptP effector tag.
29. The method of claim 28, wherein the tag comprises the sequence of SEQ ID No:1.
30. The method of claim 28, wherein the tag comprises the sequence of SEQ ID No:2.
31. The method of claim 13, wherein the expression of the protein is optimized by substituting *Salmonella* codons for insect codons in the gene expressing the protein.
32. The method of claim 13, wherein the repeat regions and mRNA structures are reduced by making use of the codon degeneracies.
33. The method of claim 1, wherein the protein is a protein that can modify a chemical/biopolymer/substrate that cannot pass through the cellular membrane of the bacterium.
34. The method of claim 1, wherein the expression of the protein is controlled by a stationary phase promoter, an inducible promoter, or a constitutive promoter.
35. The method of claim 32, wherein the stationary phase promoter is selected from the group consisting of spvA, spvR, and ssaG.
36. The method of claim 33, wherein the protein is a cellulase.

37. The method of claim 1, where the tag is from the SopE, InvJ, or SipA effectors.

38. The method of claim 1, wherein the expression of the protein is controlled by a stationary phase promoter.

39. The method of claim 1, wherein the expression of the protein is controlled by a stationary phase promoter selected from the group consisting of *spvA*, *spvR*, and *ssaG*.

40. The method of claim 1, wherein competing secretion systems of the bacterium are knocked out.

41. The method of claim 40, wherein SPI-2 and flagella in the bacterium are knocked out or inactive.

42. The method of claim 1, wherein extracellular or intracellular proteases of the bacterium are knocked out or inactive.

43. The method of claim 1, wherein the expression of the protein is under the control of hybrid promoters that are IPTG inducible.

44. The method of claim 43, wherein lac operator sites are located on either end of the promoter.

45. The method of claim 44, wherein the promoters comprise *sicA*, *spvA*, or *ssaG*.

46. The method of claim 1, wherein a *sicA* promoter drives the expression of T7 polymerase which then very strongly upregulates a T7 promoter.

47. The subject matter of any of the above claims, wherein the expression system comprises one or more genetic control element(s) which link the expression of a heterologous gene to the completion of a functional TTSS in a gram negative bacteria.

48. A fusion protein, wherein said protein is heterologous to wild-type *Salmonella*, and comprises a polypeptide tag that directs a protein to a *Salmonella* Type III secretion system.

49. The protein of claim 48, wherein the protein comprises a spider silk protein, a silk worm silk protein, or elastin.

50. The protein of claim 48, wherein the protein comprises a fibroin or a protein-based biopolymers.

51. The protein of claim 48, wherein the protein comprises a blood or plasma protein.

52. The protein of claim 48, wherein the protein comprises a mammalian peptide hormone, growth factor, cytokine, antibody, enzyme, or receptor.

53. The protein of claim 46, wherein the tag is linked to a protein of interest via a amino acid sequence subject to hydrolysis by a predetermined protease which is specific to the amino acid sequence insofar as it does not internally cleave the protein of interest.

54. The protein of claim 53, wherein the protein further comprises a polypeptide tag used in the detection or purification of the protein.

55. A polynucleotide encoding the protein of any one of claims 48 to claim 54.

56. A *Salmonella* bacterium transfected with the polynucleotide of claim 55.

57. The bacterium of claim 56, wherein the bacterium has no or reduced expression of an effector secreted by a Type III secretory system.

58. A vector or expression cassette comprising the polynucleotide of claim 55.

59. The vector or expression cassette of claim 58, wherein the polynucleotide is operably linked to a gene regulatory element that coordinates the expression of the protein encoded by the polynucleotide with a functioning Type III secretion system.

60. A kit comprising a *Salmonella* bacterium having no or reduced expression of an effector secreted by a Type III secretory system and a vector or expression cassette comprising a gene regulatory element capable of coordinating the expression of a polynucleotide with the expression of a functioning Type III secretory system in the bacterium.

61. A kit comprising:
a first container holding a *Salmonella* bacterium having no or reduced expression of an effector secreted by a Type III secretory system, and
a second container holding a polynucleotide encoding a *sicA* promoter.

62. The subject matter of any of the above claims wherein the bacterium is *Salmonella*, *Salmonella typhimurium*, or *Salmonella typhimurium* 1344.

63. The subject matter of any one of claims 1 to 62, wherein a chaperone associated with the secretion tag is co-expressed with the heterologous protein.

64. The subject matter of claim 63, wherein the chaperone and the heterologous protein are under the control of the same promoter.

65. The subject matter of claim 61, wherein the chaperone is SicP.

1/16
Human Polymers

>gi|5881413|ref|NP_000492.1| elastin [Homo sapiens]

MAGLTAAAPRPGVLLLLLSILHPSRPGGVPGAIPGGVPGGVFYYPGAGLGAL
GGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVTFPGALVPGGVADAA
AAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGLPG
VYPGGVLPGARFPGVGVLPGVPTGAGVKPKAPGVGGAFAGIPGVGPFGGP
QPGVPLGYPIKAPKLPGGYGLPYTTGKLPYGYGPGGVAGAAGKAGYPTGT
GVGPQAAAAAAAAAKAAKFGAGAAGVLPGVGGAGVPGVPGAIPGIGGIAG
VGTPAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAGVPGVG
VPGAGIPVVPAGIPGAAVPGVVSPEAAAKAAKAAKYGARPGVGVGGIP
TYGVGAGGFPGFGVGVGGIPGVAGVPSVGGVPGVGGVPGVGISPEAQAAA
AAKAAKYGVGTPAAAAAKAAKAAQFGLVPGVGVAPGVGVAPGVGVAP
GVGLAPGVGVAPGVGVAPGVGVAPGIGPGGVAAAAKSAAKVAAKAQLR
AAAGLGAGIPGLGVGVGPGLGVGAGVPGLGVGAGVPFGFAGADEGVR
RSLSPELREGDPSSSQHLPSTPSSPRVPGALAAKAAKYGAAVPGVLGGLG
ALGGVGIPGGVVGAGPAAAAAAAAAKAAKAAQFGLVGAAGLGGLGVGGL
GVPGVGGGIPPAKAAKYGAAGLGGVLGGAGQFPLGGVAARPGFG
LSPIFPGGACLGKACGRKRK

>gi|5901944|ref|NP_008977.1| elastin microfibril interfacier 1 [Homo sapiens]

MAPRTLWSCYLCLLTAAAGAASYPPRGFSLYTGSSGALSPGGPQAQIAPR
PASRHRNWCAVYVTRTVSCVLEDGVETYVKYQPCAWGQPQCPQSIMYRR
FLRPRYRVAYKTVTDMEWRCCQGYGGDDCAESPAPALGPASSTPRPLARP
ARNLSGSSAGSPLSGLGGEGPGESEKVQQLEEQVQSLTKELQGLRGVLQG
LSGRLAEDVQRAVETA FNGRQQPADAAARPGVHETLNEIQHQLQLLDTRV
STHDQELGHLNNHHGGSSSSGSRAPAPASAPPGPSEELLRQLEQRLQESC
SVCLAGLDGFRRQQQEDRERLRAMEKLLASVEERQRHLAGLAVGRPPQE
CCSPELGRRLAELERRLDVVAGSVTVLSGRRGTELGAAGQGHPGPGYTS
LASRLSRLED RFNSTLGPSEEQEE SWPGAPGGLSHWLPAARGRLEQLGGLL
ANVSGELGGRLDLLEE QVAGAMQACGQLCSGAPGEQDSQVSEILSALERR
VLDSEGQLRLVGSGLHTVEAAGEARQATLEGLQE VVGRLQDRVDAQDET
AAEFTLRLNLTAARLGQLEGLLQAHGDEGCGACGGVQEELGRLRDGVER
CSCPLLPPRPGPAGPGVGGPSRGPLDGFSVFGSSGSALQALQGELSEVILS
FSSLNDSLNELQTTVEGQGADLADLGATKDRIISEINRLQQEATEHATESEE
RFRGLEEGQAQAGQCPSLEGRLGRLEGVCERLDTVAGGLQGLREGLSRHV
AGLWAGLRETNTTSQMQAALLEKL VGGQAGLGRRLGALNSSLQLLEDRL
HQLSLKDLTG PAGEAGPPGPPGLQGPPGPAGPPGSPGKDGQEGPIGPPGPQ
GEQGVGAPAAPVPQVAFSAALSLPRSEPGTVPFDRVLLNDGGYYDPETG
VFTAPLAGRYLLSAVLTGHRHEKVEAVLSRSNQGVARVDSGGYEPEGLEN
KPVAESQSPGTLGVFSLILPLQAGDTV CVDLVMGQLAHSEEPLTIFSGALL
YGDPELEHA

FIG. 1

Secretion Signal Sequences

SEQ ID NO:1 SptP secretion signal

MLKYEERKLNNLTLSSFSKVGVSNDARLYIAKENTDKAYVAPEKFSSK
VLTWLGKMPLFKNTEVVQKHTENIRVQDQKILQTFLHALTEKYGETAV
NDALLMSRINMNKPLTQRLAVQITECVKAADEGFNLIKSKDNVGVNRN
AALVIKGGDTKVAEKNNDVGAES

SEQ ID NO:2 InvJ secretion signal

MGDVSAVQ

SEQ ID NO:3 SopE secretion signal 100:

MTNITLSTQHYRIHRSDVEPVKEKTTEKDIFAKSITAVRNSFISLSTSLSD
RFSLHQQTDIPTTHFHRGNASEGRAVLTSKTVKDFMLQKLNSLDIKGN

FIG. 2

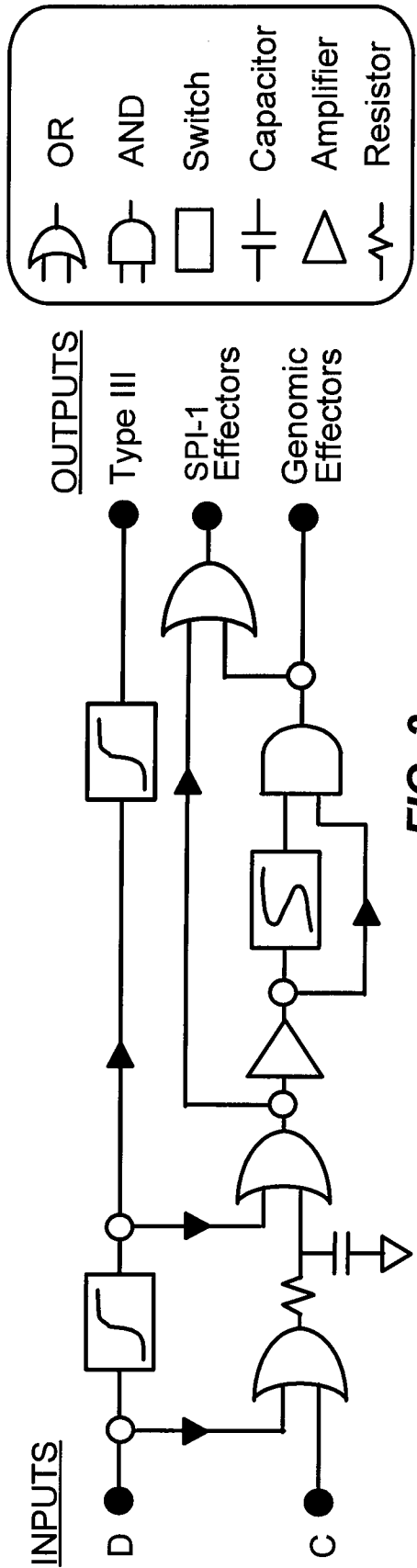
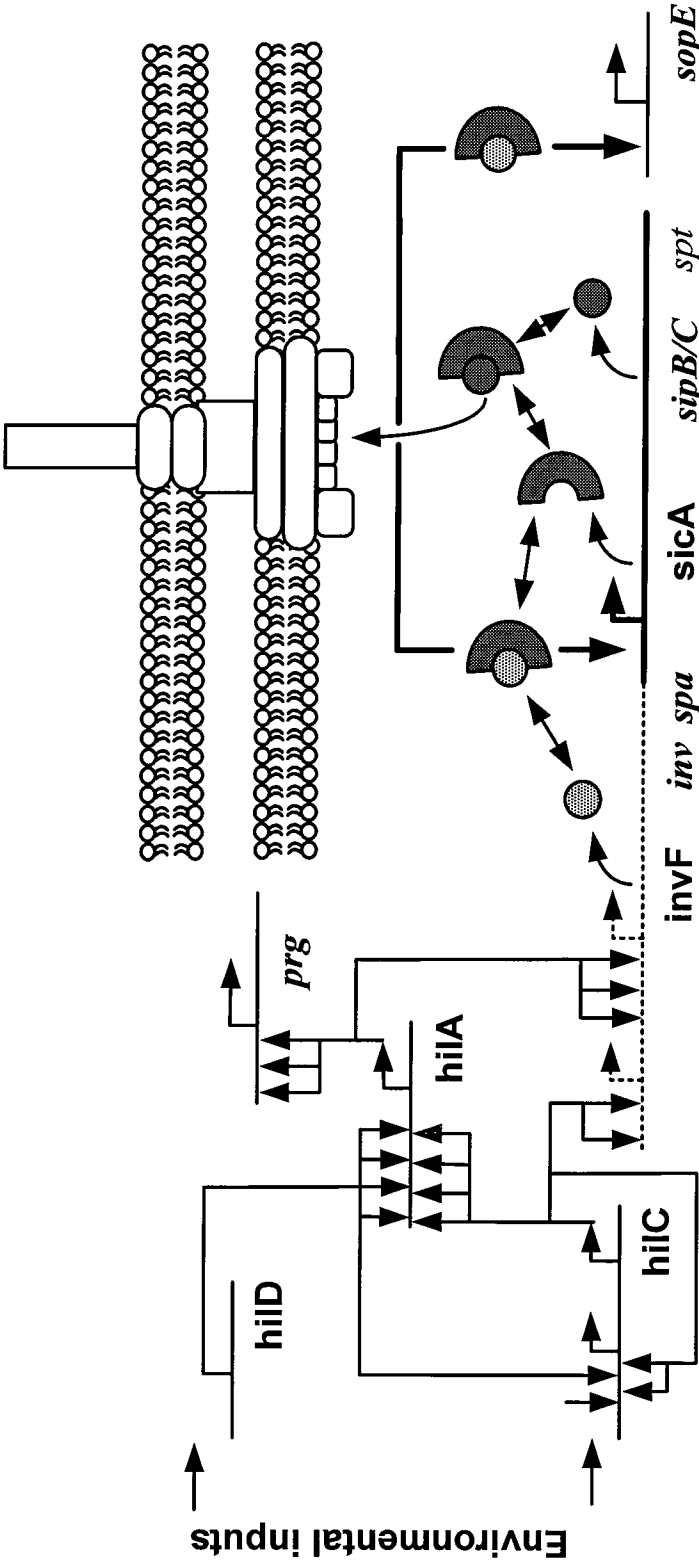


FIG. 3

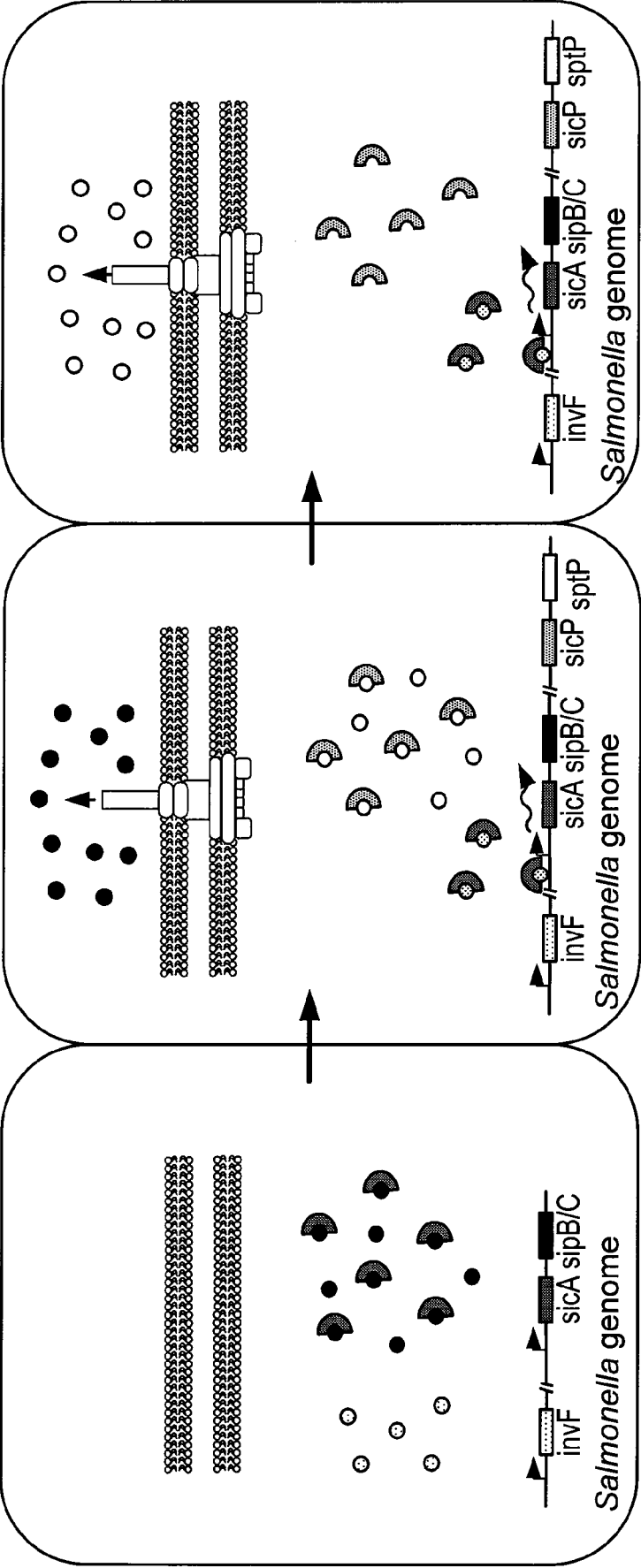
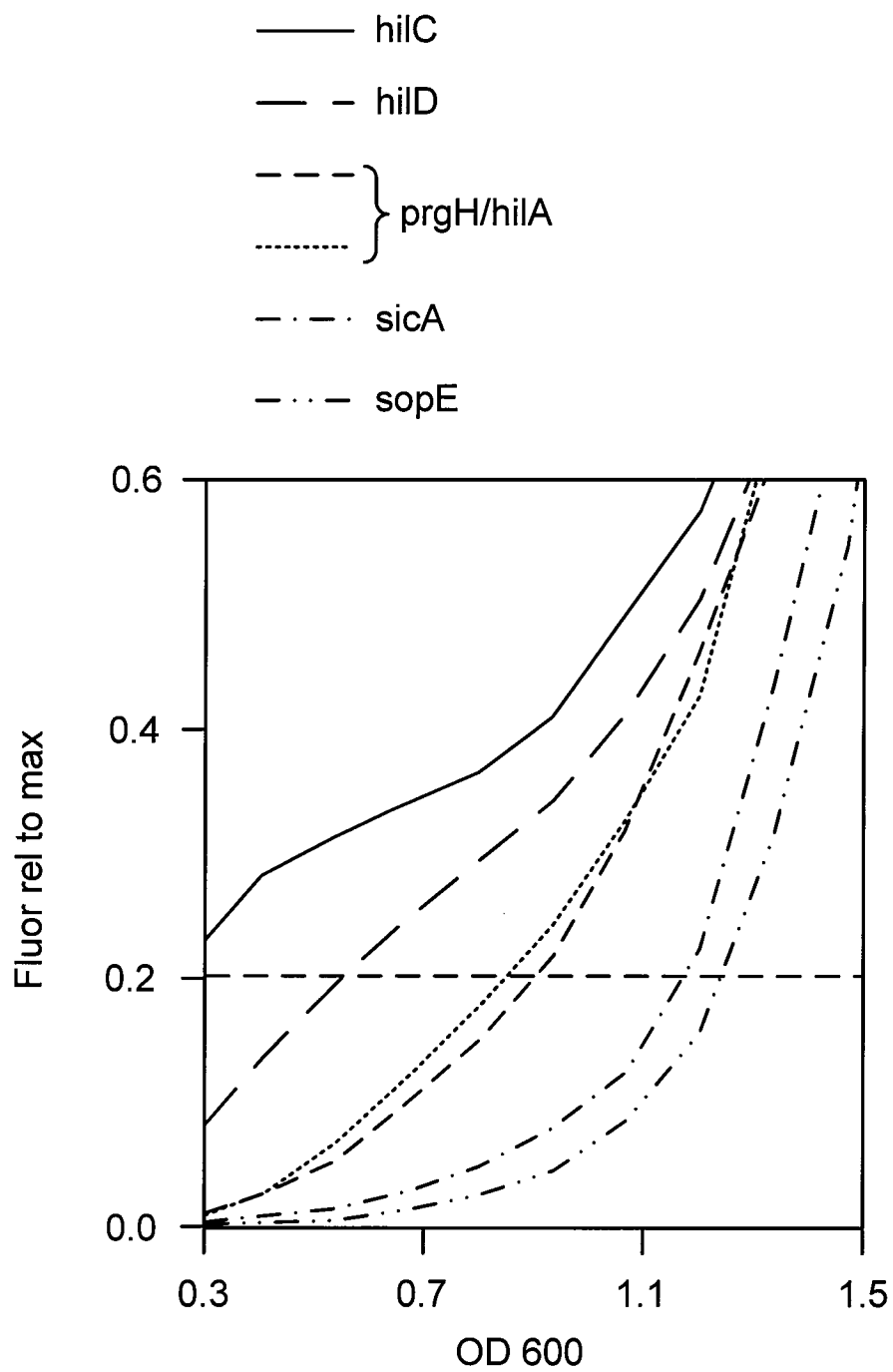
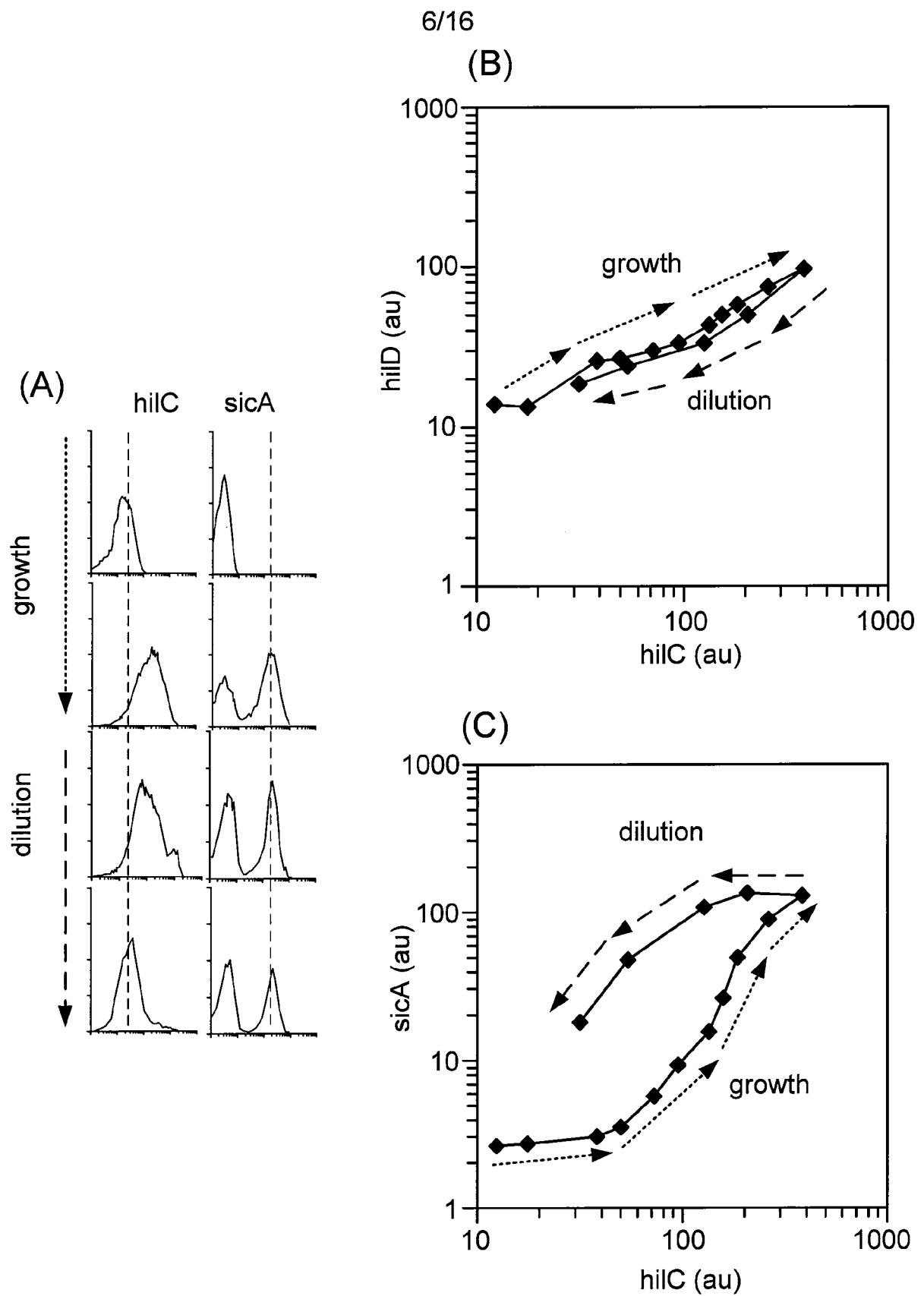


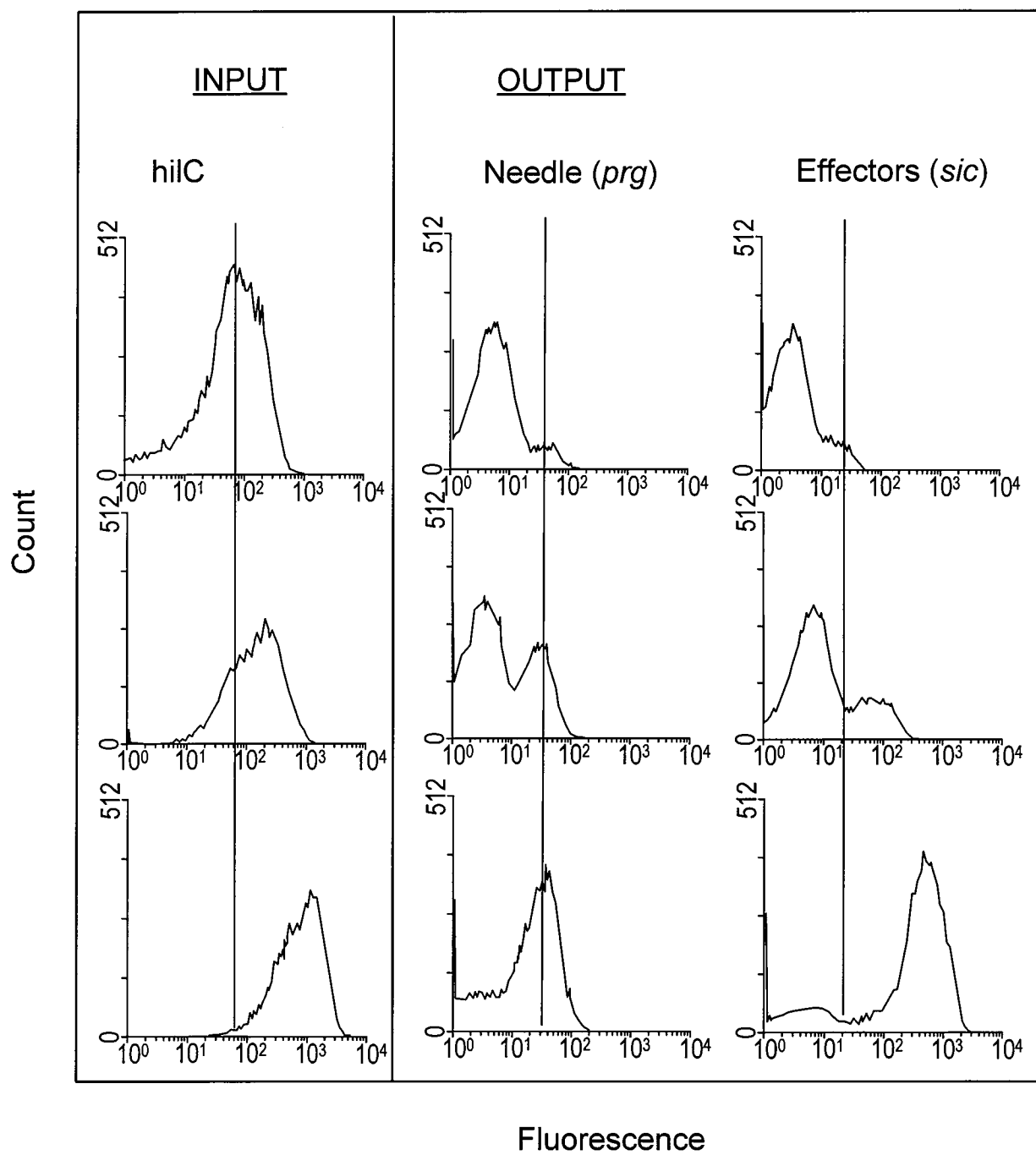
FIG. 4

5/16

**FIG. 5**

**FIG. 6**

7/16

**FIG. 7**

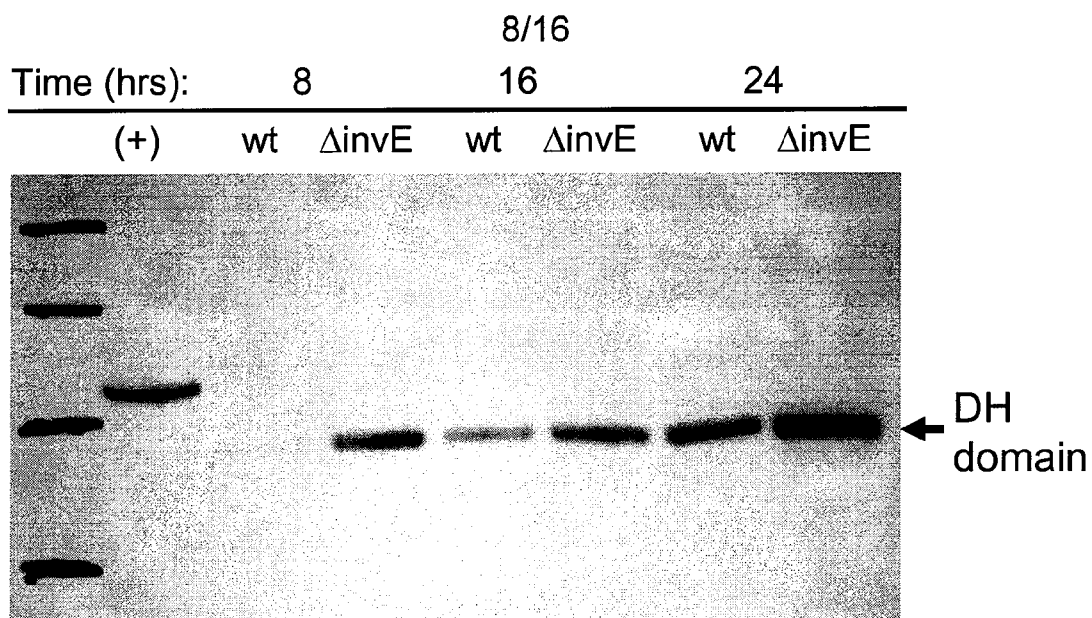


FIG. 8

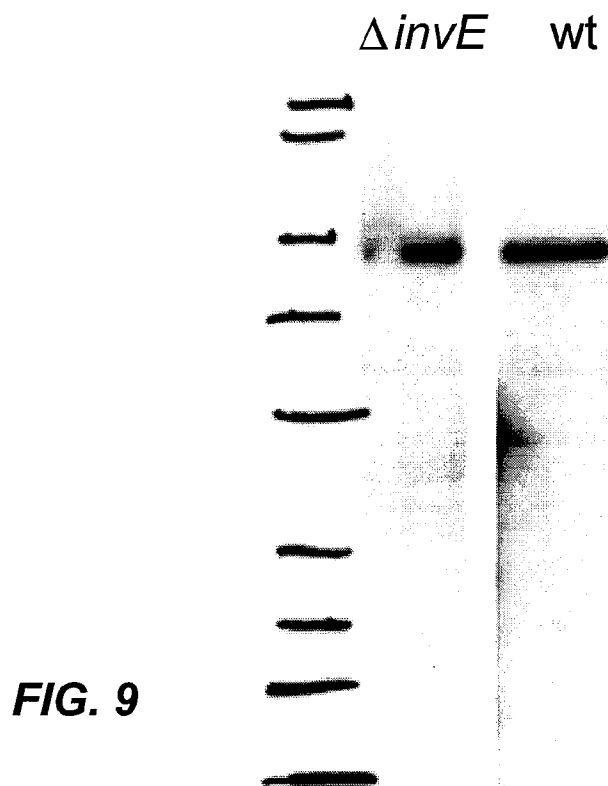


FIG. 9

FLAG tag: DYKDDDDKDYKDDDDKDYKDDDDK

FIG. 10

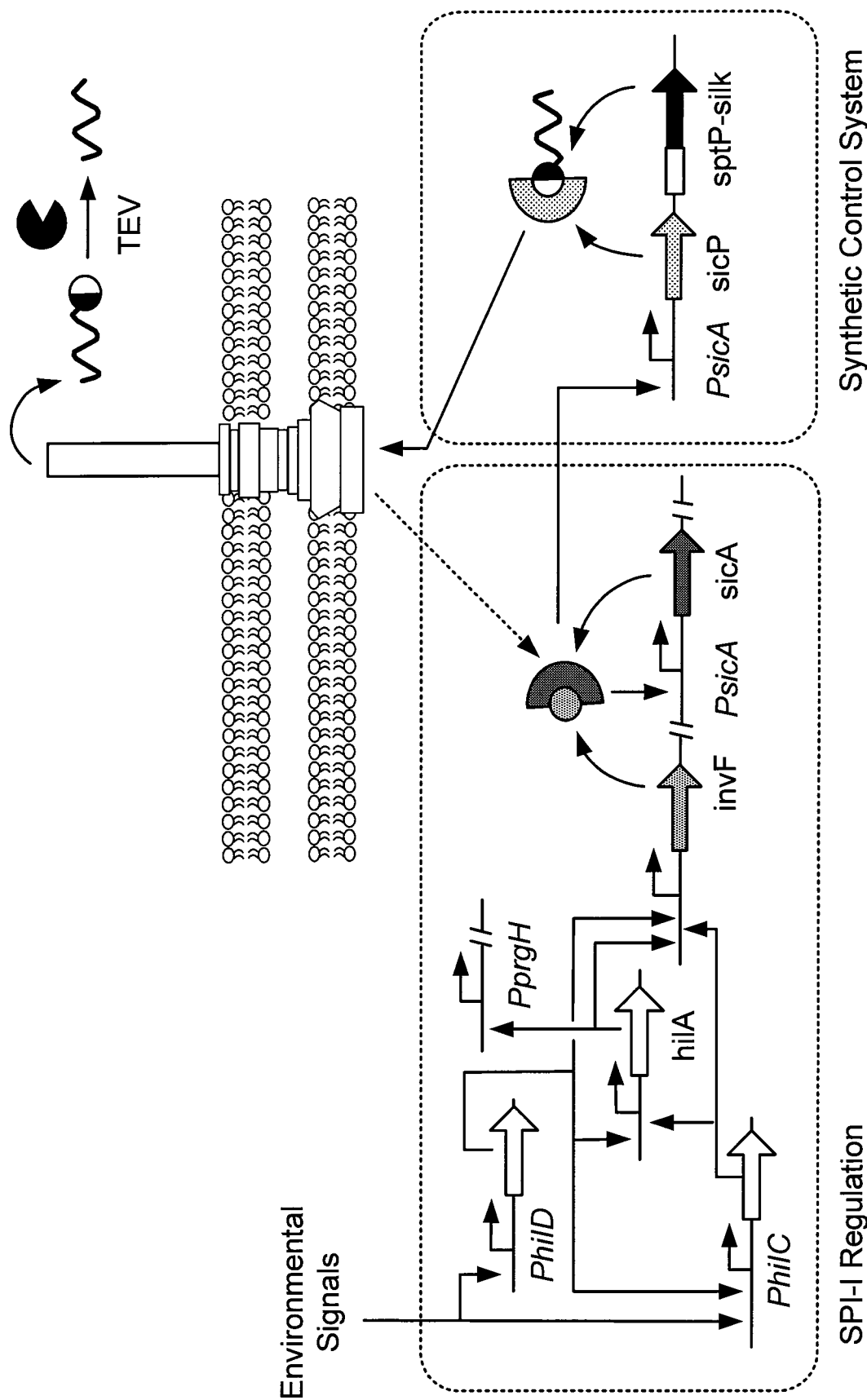


FIG. 11

10/16

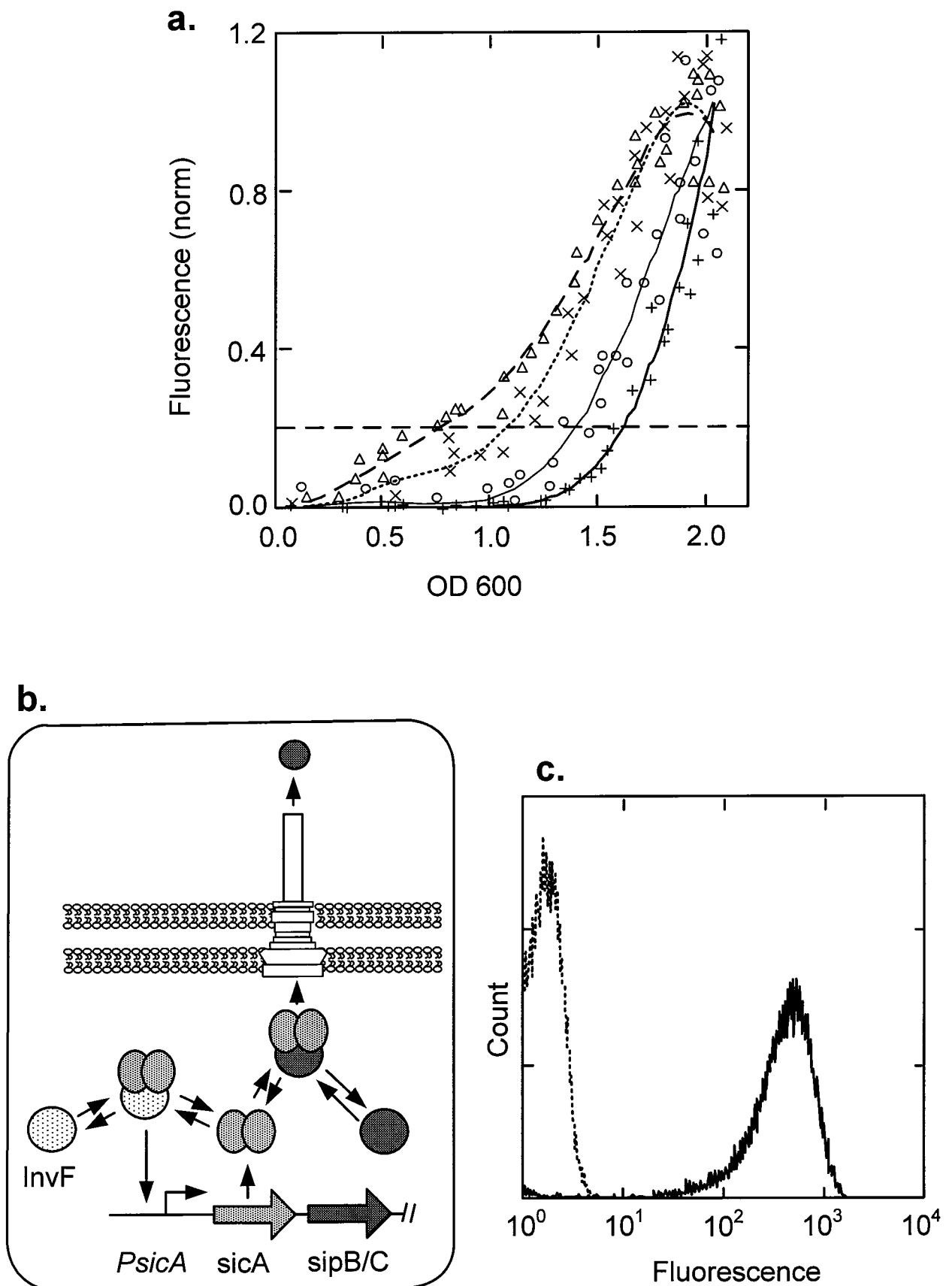


FIG. 12

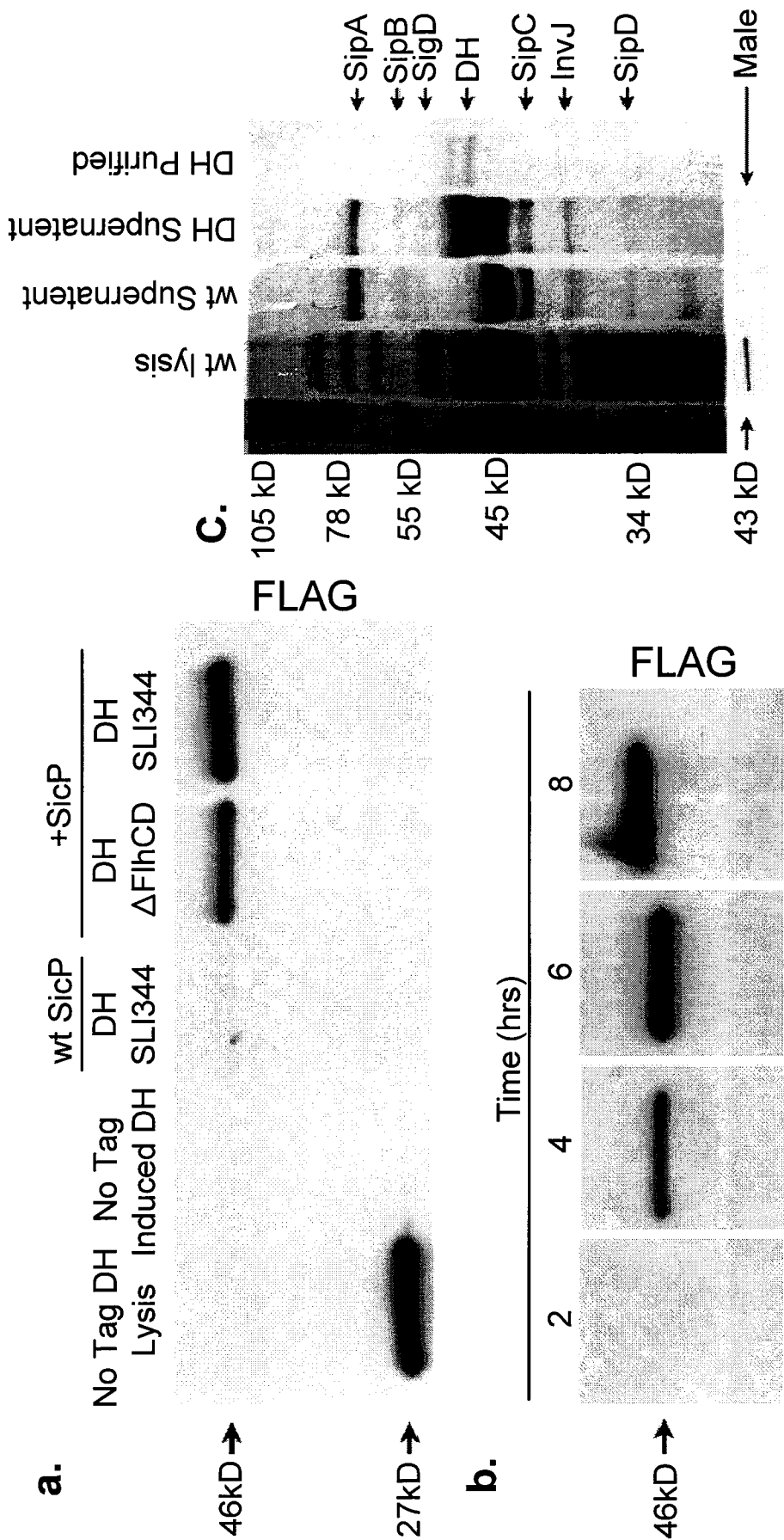


FIG. 13

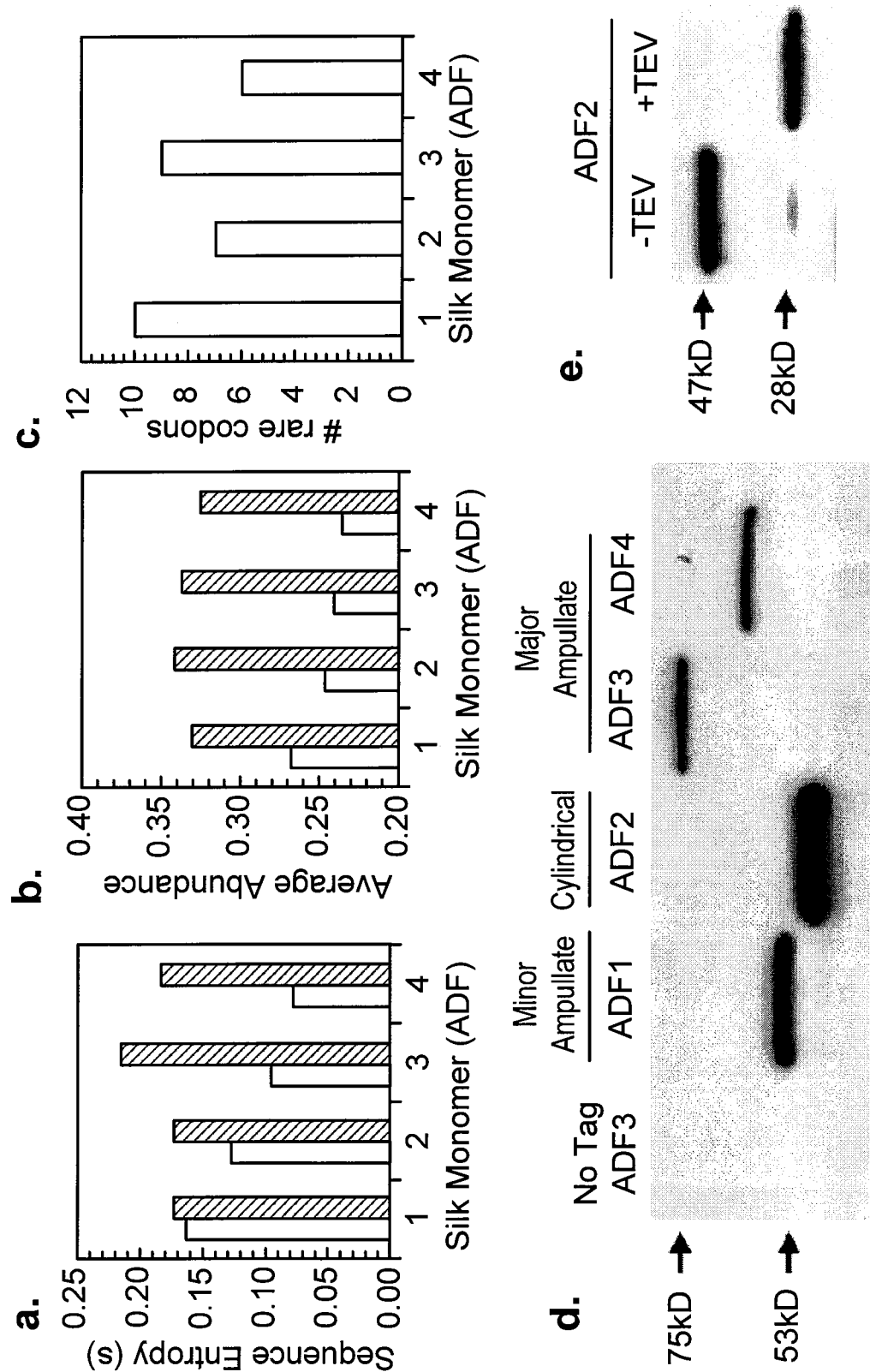


FIG. 14

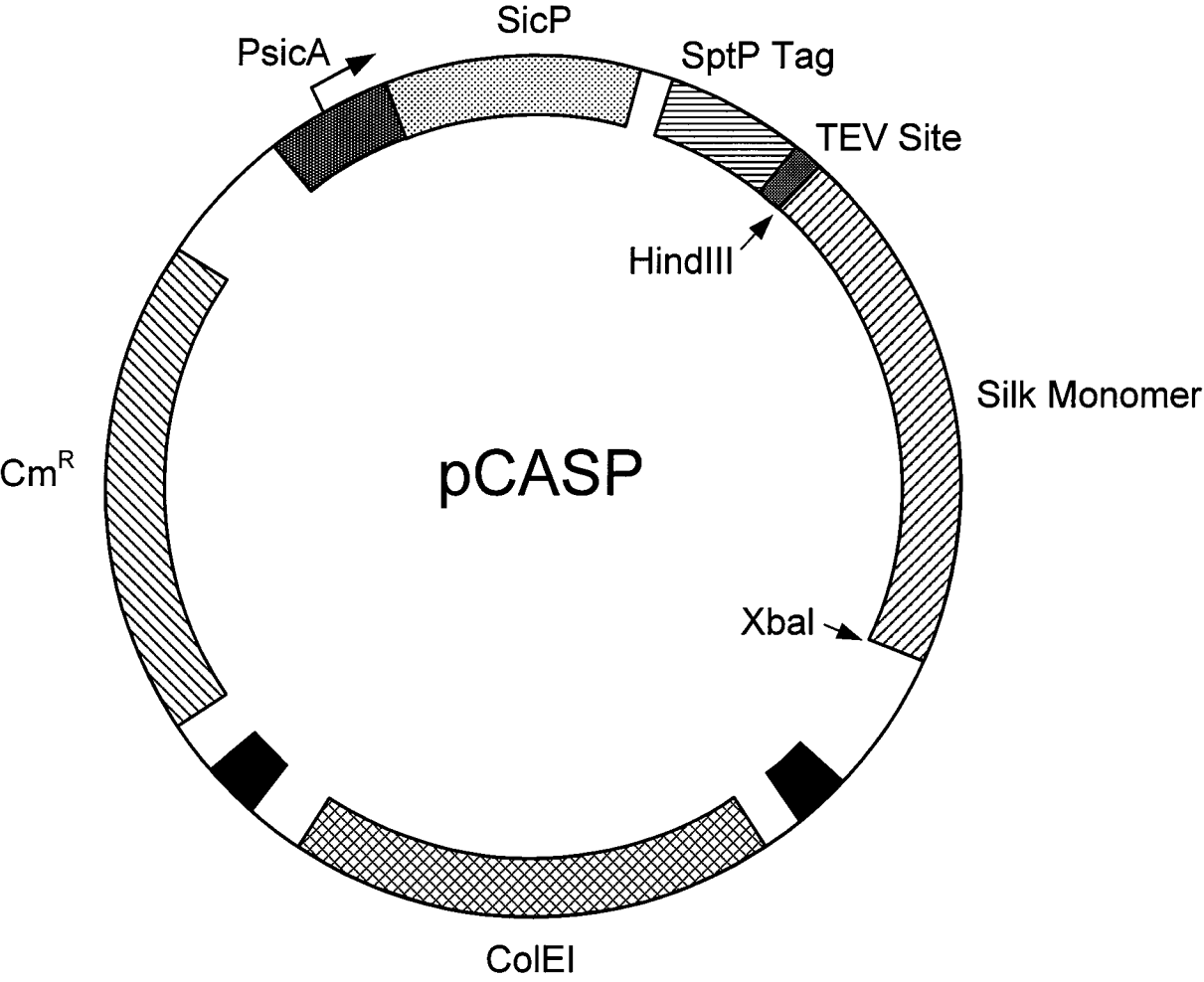


FIG. 15

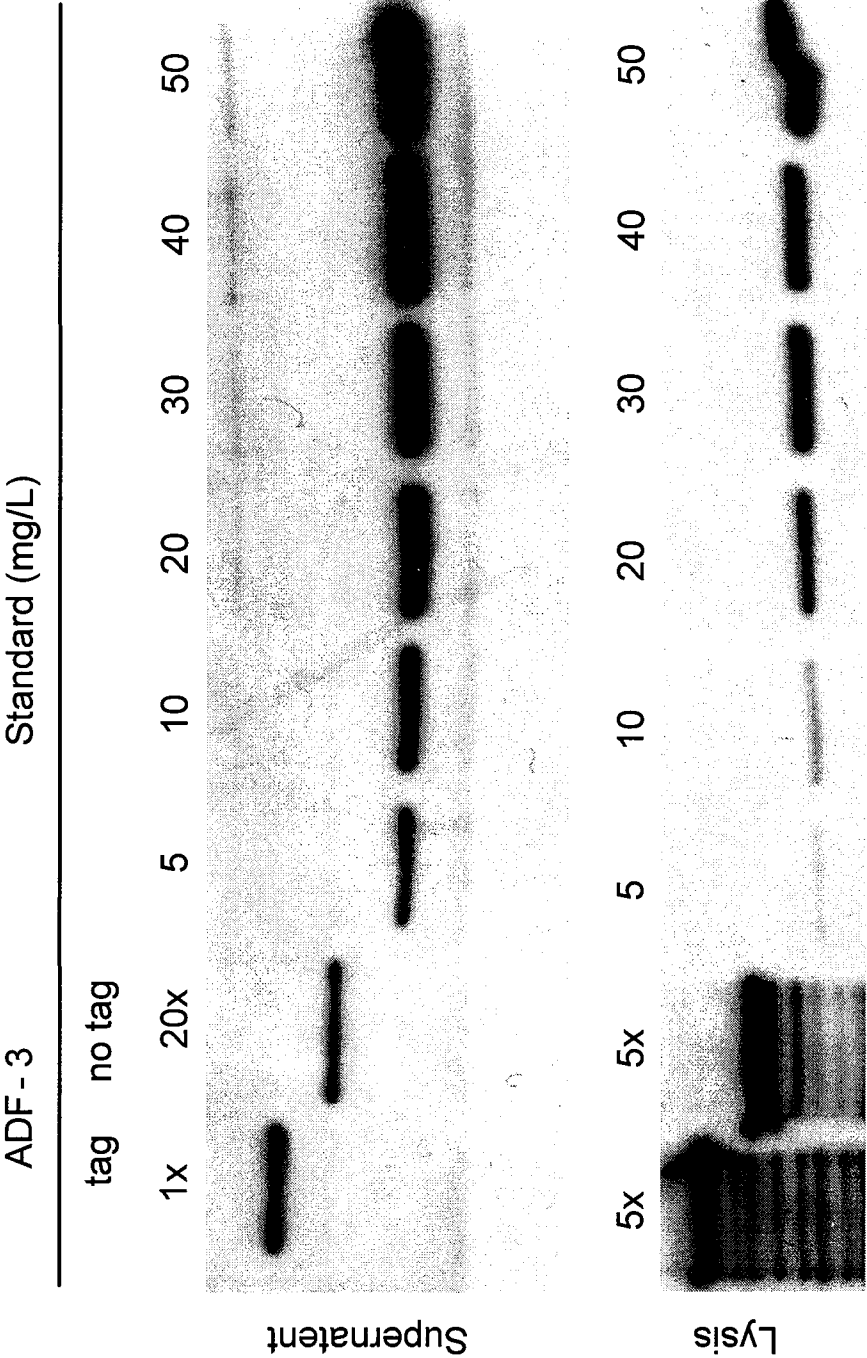


FIG. 16

Comparison of secretion signal, chaperone binding domain and chaperone

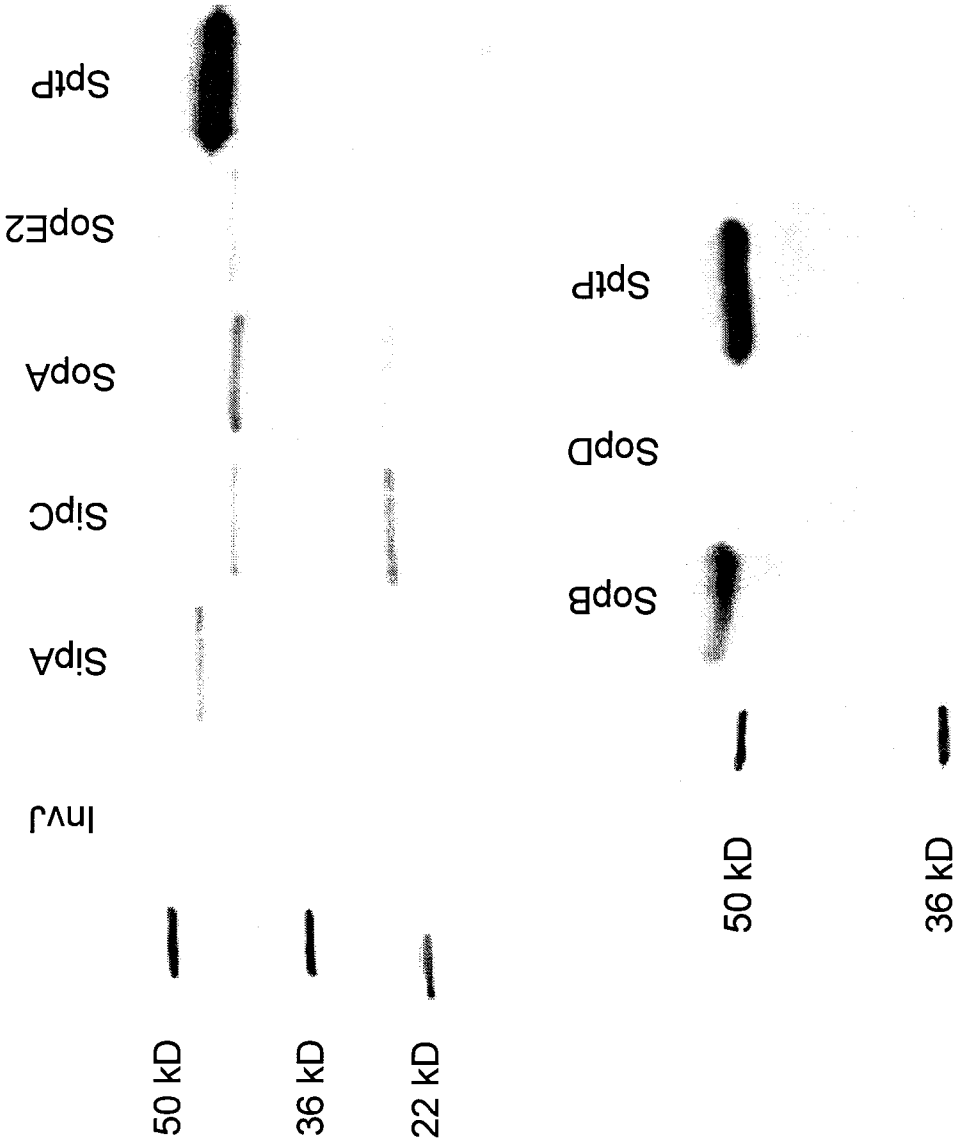


FIG. 17

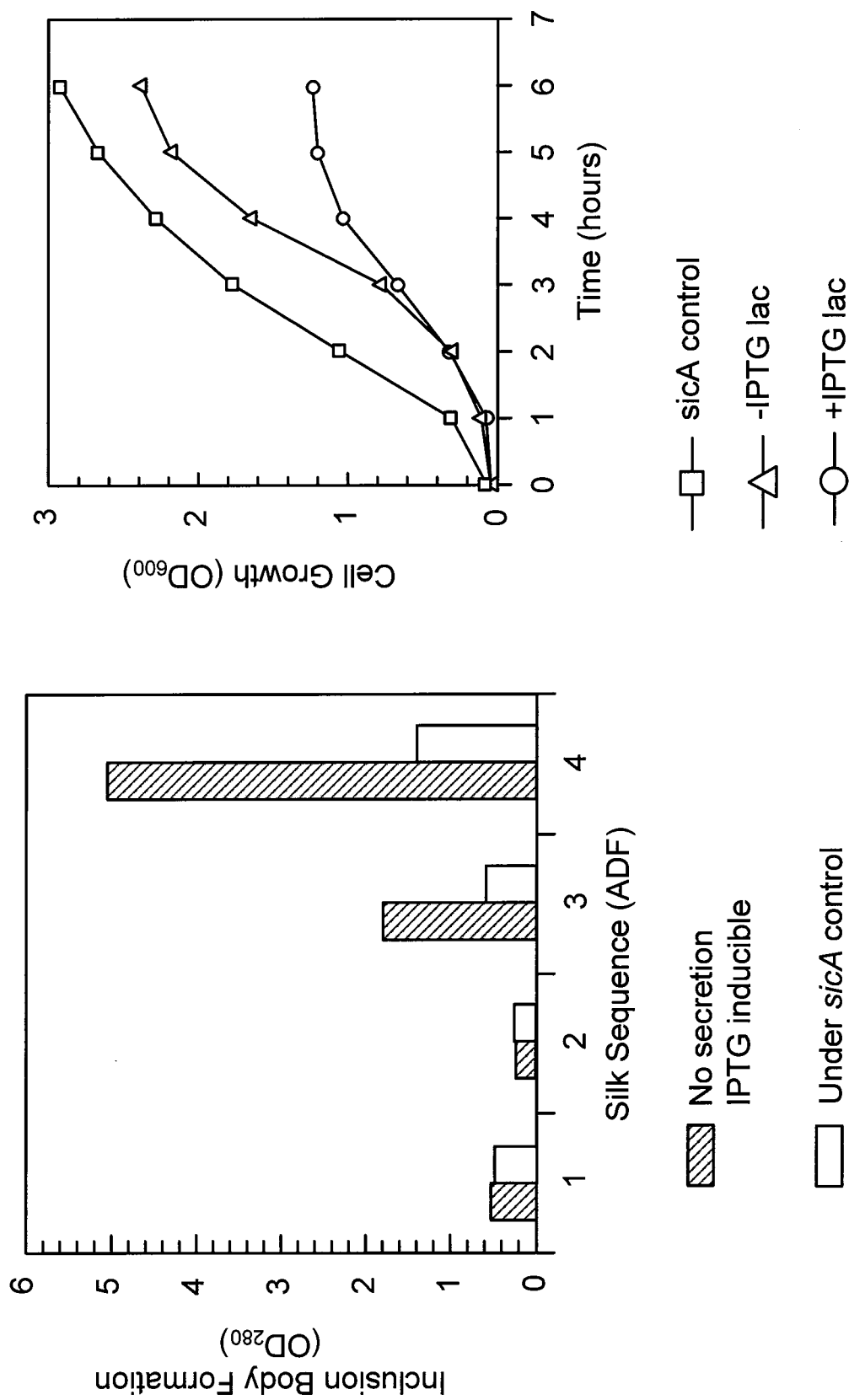


FIG. 18