Molecular Systems Biology 5; Article number 309; doi:10.1038/msb.2009.62
Citation: ***Molecular Systems Biology*** **5:309**

www.molecularsystemsbiology.com

molecular systems biology

REPORT

Engineering the *Salmonella* type III secretion system to export spider silk monomers

Daniel M Widmaier[1], Danielle Tullman-Ercek[2], Ethan A Mirsky[3], Rena Hill[2], Sridhar Govindarajan[4], Jeremy Minshull[4] and Christopher A Voigt[1,2,3,*]

[1] Chemistry and Chemical Biology Graduate Program, University of California—San Francisco, San Francisco, CA, USA, [2] Department of Pharmaceutical Chemistry, University of California—San Francisco, San Francisco, CA, USA, [3] Graduate Group in Biophysics, University of California—San Francisco, San Francisco, CA, USA and [4] DNA2.0, Menlo Park, CA, USA

* Corresponding author. Department of Pharmaceutical Chemistry, University of California—San Francisco, Box 2540, Room 408C, 1700 4th Street, San Francisco, CA 94110, USA. Tel.: +1 415 5027 050; Fax: +1 415 5024 690; E-mail: cavoigt@picasso.ucsf.edu

Received 5.5.09; accepted 24.7.09

The type III secretion system (T3SS) exports proteins from the cytoplasm, through both the inner and outer membranes, to the external environment. Here, a system is constructed to harness the T3SS encoded within *Salmonella* Pathogeneity Island 1 to export proteins of biotechnological interest. The system is composed of an operon containing the target protein fused to an N-terminal secretion tag and its cognate chaperone. Transcription is controlled by a genetic circuit that only turns on when the cell is actively secreting protein. The system is refined using a small human protein (DH domain) and demonstrated by exporting three silk monomers (ADF-1, -2, and -3), representative of different types of spider silk. Synthetic genes encoding silk monomers were designed to enhance genetic stability and codon usage, constructed by automated DNA synthesis, and cloned into the secretion control system. Secretion rates up to 1.8 mg l^{-1} h^{-1} are demonstrated with up to 14% of expressed protein secreted. This work introduces new parts to control protein secretion in Gram-negative bacteria, which will be broadly applicable to problems in biotechnology.
Molecular Systems Biology **5:** 309; published online 15 September 2009; doi:10.1038/msb.2009.62
Subject Categories: synthetic biology
Keywords: automated DNA synthesis; cellular engineering; biomaterials; genetic parts; systems biology

Introduction

A difficult problem in engineering Gram-negative bacteria is the directed secretion of proteins to the extracellular environment (Wittrup, 2000; Harvey *et al*, 2004; Lee *et al*, 2006; Zhang *et al*, 2006). There are few secretion systems that are capable of exporting proteins through both the inner and outer membrane. This is an important tool for several applications in biotechnology. For example, the expression of some recombinant proteins at high titers can lead to the formation of inclusion bodies or retard cell growth (Sorensen and Mortensen, 2005). Secretion is also critical when the function of the protein requires that it be outside of the cell, as is the case for cellulases and other polymer-degrading enzymes that act on substrates that cannot cross the cell membrane. Here, we have harnessed the type III secretion system (T3SS) encoded on *Salmonella* Pathogeneity Island 1 (SPI-1) as a tool to export proteins of biotechnological interest. This is applied to the recombinant production of spider silk proteins, which can form fibrils if they are allowed to accumulate inside of the confined volume of the cell (Huemmerich *et al*, 2004a).

The T3SS is unique because it a well-characterized protein secretion system that 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 (Georgiou and Segatori, 2005; Wickner and Schekman, 2005). The Sec pathway threads polypeptides in an unfolded state across the inner membrane in an ATP-dependent manner (Economou and Wickner, 1994; Pohlschroder *et al*, 2005). The Tat pathway uses the proton motive force to drive the transport of folded proteins to the periplasm (Sargent *et al*, 1998; Rodrigue *et al*, 1999; DeLisa *et al*, 2003). An N-terminal signal peptide is required for both Sec and Tat export (Berks, 1996; Wickner and Schekman, 2005). Both types of signal peptides typically end with a signal peptidase I cleavage site, allowing cleavage of the tag on translocation to the periplasm (Nielsen

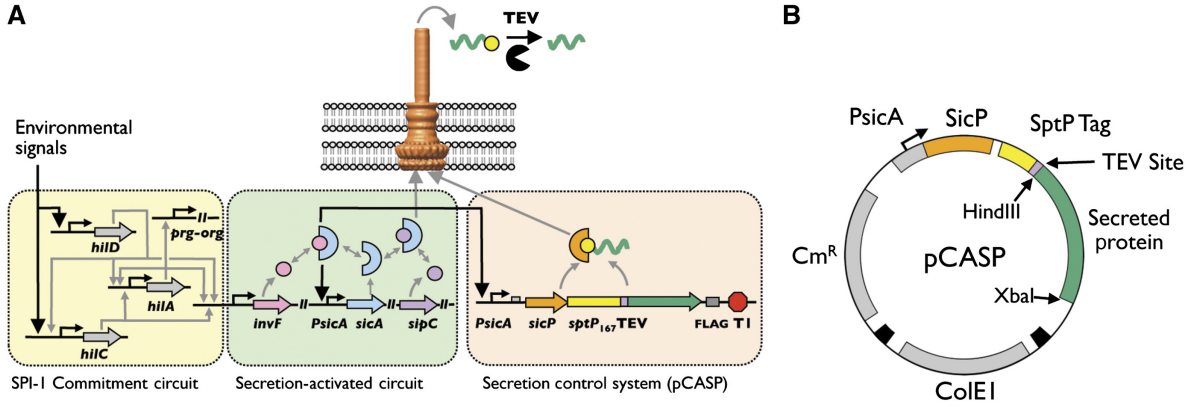
et al, 1997; Bendtsen *et al*, 2005). Type II secretion can export proteins from the periplasm through the outer membrane; however, the secretion signal is difficult to identify and seems to be distributed throughout the protein, making heterologous protein secretion difficult (Pohlschroder *et al*, 2005). Remarkably, it has been shown that all of the genes associated with the *Erwinia* type II secretion system can be transferred to *Escherichia coli* and used to secrete *Erwinia* cellulases (Zhou *et al*, 1999). Alternatively, Gram-positive bacteria offer a single secretion event to the extracellular space and offer an attractive platform for secretion engineering.

In its natural context, the SPI-1 T3SS functions as a molecular syringe to inject effector proteins into mammalian host cells that facilitate invasion and pathogenesis (Altier, 2005). The SPI-1 T3SS forms a needle-like structure that crosses both the inner and outer membranes (Marlovits *et al*, 2006). A chaperone is required for secretion, as well as an N-terminal peptide tag that is not cleaved post secretion (Galan and Collmer, 1999). Secreted proteins can be folded in the cytoplasm and then are unfolded in an ATP-dependent reaction before secretion (Feldman *et al*, 2002; Lee and Schneewind, 2002; Akeda and Galan, 2005). It is expected based on needle dimensions that proteins must be at least partially unfolded to transit through the needle and would be required to re-fold outside the cell. The *E. coli* and *Salmonella* flagellum and *Yersinia enterocolitica ysc* T3SS have been shown to be able to export heterologous proteins (Russmann *et al*, 1998; Feldman *et al*, 2002; Lee and Schneewind, 2002; Majander *et al*, 2005; Chen *et al*, 2006; Konjufca *et al*, 2006; Vegh *et al*, 2006). These systems have been used to inject foreign proteins and peptides into mammalian cells as a mechanism to confer immunity (Russmann *et al*, 1998; Boyd *et al*, 2000; Konjufca *et al*, 2006).

A well-characterized regulatory network encoded within SPI-1 controls the dynamics of T3SS gene expression (Box 1) (Lucas and Lee, 2000). Environmental signals from two-component systems and global regulators control the expression of the HilC, HilD, and HilA transcription factors, which together form a commitment circuit for the expression of SPI-1 genes (Bajaj *et al*, 1996; Eichelberg and Galan, 1999; Lundberg *et al*, 1999; Kalir *et al*, 2001; Ellermeier *et al*, 2005). Within SPI-1, there is a genetic circuit that links the expression of effector proteins to the completion of functional needles (Darwin and Miller, 1999, 2000, 2001; Temme *et al*, 2008). The circuit consists of a transcription factor (InvF) that is only functional when bound to the SicA chaperone protein. Before the cell can secrete protein, the chaperone is sequestered by the SipB/C proteins. After functional needles are completed, SipB/C are secreted and SicA is free to bind InvF, thus turning on the circuit and gene expression from the *sicA* promoter (Darwin and Miller, 1999, 2000, 2001; Tucker and Galan, 2000).

The *Salmonella* SPI-1 T3SS has several properties that make it a good tool for the secretion of recombinant proteins. It is highly expressed under standard laboratory conditions (Luria-Bertani Broth at 37°C), with 10–100 needles per cell (Kubori *et al*, 1998). Under these conditions, effector proteins are secreted into the media in significant quantities without the need to co-culture with mammalian cells or expensive media components (Kubori and Galan, 2002). Finally, the N-terminal secretion tags, chaperone-binding domains (CBD), and chaperones have been identified (Fu and Galan, 1998; Hong and Miller, 1998; Bronstein *et al*, 2000; Tucker and Galan, 2000; Russmann *et al*, 2002; Zhang *et al*, 2002; Ehrbar *et al*, 2003; Lee and Galan, 2004; Wood *et al*, 2004; Karavolos *et al*, 2005; Higashide and Zhou, 2006; Knodler *et al*, 2006). On the basis of

Box 1 A genetic circuit from the SPI-1 regulatory pathway is harnessed to control expression



Box 1 The regulatory network controlling T3SS self-assembly is shown (yellow background) (panel **A**). Within this network, an operon containing the chaperones, translocators, and effectors is controlled by a genetic circuit that becomes active once the T3SS is constructed and functional (green background, centered on *invF* and *sicA*). A secretion control system is constructed that contains all of the necessary parts for the T3SS to export heterologous proteins (orange background). The *sicA* promoter and ribosome-binding site drive the expression of the chaperone (SicP) and heterologous protein (green) fused to an N-terminal secretion signal (SptP). A TEV protease site is included after the tag such that it can be removed post secretion. The cryo-EM image of the T3SS is reproduced from reference (Marlovits *et al*, 2004). A map of the pCASP plasmid is shown (**B**). The secretion control system superpart is BBa_J64032 in the Registry of Standard Biological Parts (www.partsregistry.org) and the full pCASP sequence is available in Genbank (#EF179157).

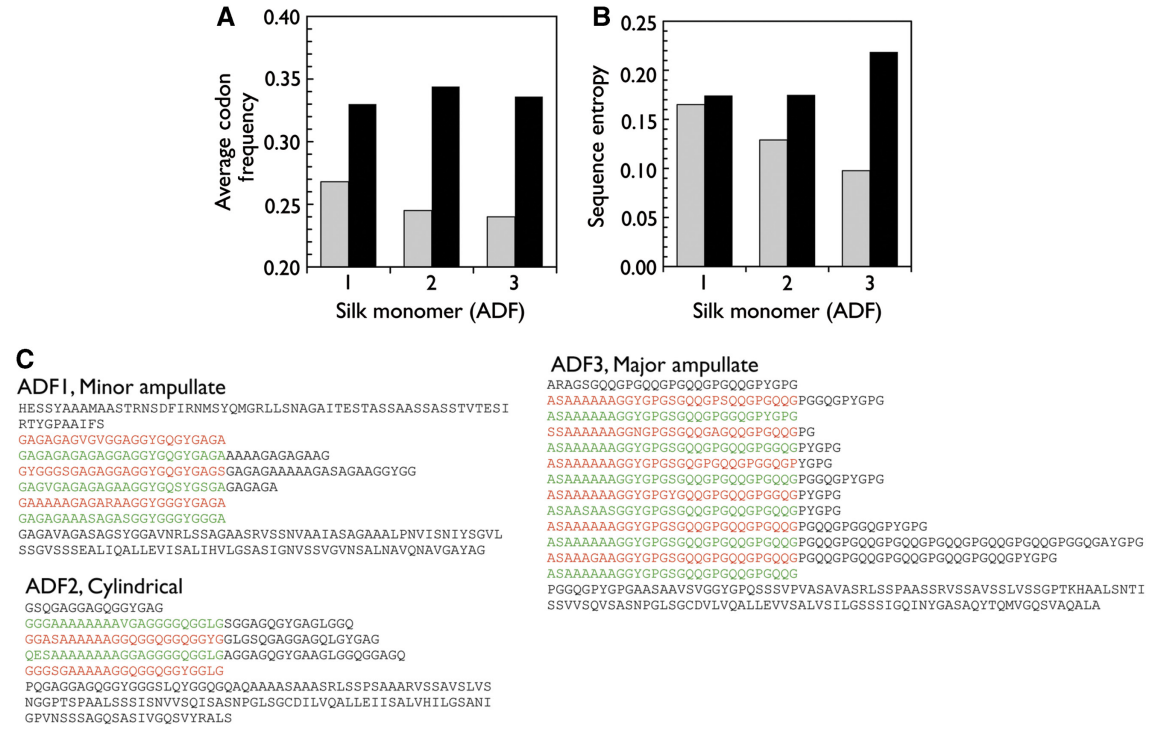


Figure 1 Changes made to the *Araneus* spider silk DNA sequences. (**A**) The codon frequency is a measure of the abundance of codon sequences in the *E. Coli* genome relative to each amino acid. The average frequency is the mean of the codon frequencies across the entire sequence of the silk gene. The average codon frequency is shown for the spider (gray) and synthetic (black) genes. Very rare codons (<10 per gene, defined as frequencies <0.13) were entirely eliminated from the sequences. (**B**) The DNA sequence entropy of the repetitive units is shown for the wild-type spider (gray) and synthetic (black) genes. The repeat units for each silk monomer were manually aligned (Supplementary information) and the sequence entropy is calculated: $S = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{4} p_i(j)\mathrm{In}p_i(j)$, 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 based are equally represented at each position) is ~ 0.6. A lower sequence entropy indicates a higher degree of sequence identity between the repeat units. ADF-3 has extremely repetitive DNA sequences and this repetitiveness is effectively eliminated upon optimization. (**C**) The amino-acid sequences of the synthetic spider silk genes are shown. Each silk sequence is labelled with a name and the gland in which it is produced. The repetitive regions used in the sequence entropy calculations are in red and green.

this earlier work, we have constructed a system that contains all of the necessary genetic parts to secrete heterologous proteins (Box 1). This system is demonstrated by using it to export a human protein and three spider silk monomers.

Natural silks are abundant biomaterials that span a remarkable diversity of physical properties (Vollrath and Knight, 2001; Swanson *et al*, 2006). For example, the dragline silk of spider webs is extremely strong yet remains highly elastic. These materials have a number of uses including medical device implants, high strength fibers, advanced composites, and drug delivery systems (Lewis, 2006; Wang *et al*, 2006a, b; Hofmann *et al*, 2007; Lee *et al*, 2009). Spiders are not conducive to scalable agriculture, therefore requiring silk production to be done in a recombinant host (Lewis, 2006). Using a solution of natural or recombinant silk monomers, it has been shown that threads can be artificially spun, producing materials with properties approaching natural silks (Seidel *et al*, 2000; Lazaris *et al*, 2002). One of the key limitations in creating materials that match or exceed the natural properties is the lack of practical approaches for recombinant protein expression (Kluge *et al*, 2008).

The production of recombinant native silk proteins has been complicated by several factors. First, the genes themselves are often unstable because of highly repetitive regions of DNA that results in frequent homologous recombination (Arcidiacono *et al*, 1998). Second, the codon usage in silk genes is not optimized for expression outside of the specialized cells in the silk gland (Prince *et al*, 1995; Rising *et al*, 2005). Rare codons result in ribosome pausing and early truncation producing incomplete protein products (Fahnestock and Irwin, 1997). Third, if the proteins are highly expressed in the confined cell volume they can self-assemble to produce fibrils, which have been visualized when recombinantly expressed in insect cells (Huemmerich *et al*, 2004b).

In this paper, we use whole gene DNA synthesis to construct long, computationally designed DNA sequences that exactly match the wild-type amino-acid sequence for the known fragments of silk monomers (Figure 1). Synthetic genes were designed and constructed for three silk genes from the orb weaving spider *Araneus diadematus*. These genes are expressed in different silk glands and vary in their amino-acid content and material properties (Vollrath and Knight, 2001). ADF-1 is expressed in the minor ampullate gland and is used during web construction, the resulting fibers have high tensile strength, but are inelastic (Guerette *et al*, 1996; Gosline *et al*, 1999). ADF-2 is expressed in the cylindrical gland (used for egg sacks) and has a sequence that is similar to human elastins. ADF-3 is expressed in the major ampullate gland and forms the

extremely tough and elastic dragline, which anchors the web and is used as a lifeline for escape. Each of the synthetic genes is expressed and exported from the cell using the *Salmonella* SPI-1 T3SS.

Results

Three spider silk genes (ADF-1, -2 and -3) were computationally designed for expression in eubacteria and constructed using automated DNA synthesis and assembly. The wild-type DNA sequences of the spider silk genes contain rare codons (Supplementary information), which can result in truncated proteins and poor expression (Gustafsson *et al*, 2004). The sequences also consist of repeated amino-acid units that correspond to repetitive underlying DNA sequences. The variation in the repeat amino-acid units is one of the key features that differentiate natural spider silks (Figure 1C). These regions can be very large with over 100 nucleotides of exact identity, which is a potential target for homologous recombination, resulting in genetic instability. The degeneracy of the genetic code allows many alternative nucleotide sequences to encode the same protein. This enables synthetic genes to be designed that simultaneously redistribute the codon usage, while reducing the DNA repetitiveness (Figure 1A and B). Each re-designed gene was constructed using automated whole gene DNA synthesis (Supplementary information).

We constructed a genetic system to secrete heterologous proteins into the culture media (Box 1). This system consists of four genetic parts. A circuit is harnessed from the natural SPI-1 regulatory pathway that controls gene expression based on the capacity of the cell to export protein (Darwin and Miller, 1999, 2000, 2001; Temme *et al*, 2008). The output of this circuit (*PsicA*) drives the transcription of an operon containing the heterologous protein fused to an N-terminal secretion tag and the associated chaperone. The human DH domain (Hussain *et al*, 2001) was chosen as an initial target because it is small (24 kDa), easy to manipulate, and it expresses well in *Salmonella* without affecting cell growth. A protease cleavage site is included between the tag and exported protein so that the tag can be removed after secretion. All of these parts are combined onto a plasmid (Box 1) and are available in the Registry of Standard Biological Parts (http://www.partsregistry.org/). The part numbers (BBa_) are provided throughout the text.

Using the SicA:InvF circuit to control expression has two advantages. First, it restricts expression until the T3SS is built and is functional (Darwin and Miller, 1999, 2000, 2001). This links expression to secretion capacity and can reduce the build-up of protein before secretion. The second advantage is that no exogenous inducer is required. The circuit can be maintained in the off state when cells are grown in low salt L Broth (LB) and then turned on after the cells are induced by high salt LB (inducing medium) and the needle is constructed (Figure 2A) (Materials and methods) (Tartera and Metcalf, 1993). The *sicA* promoter has a low basal transcription rate and increases 200-fold in activity when induced. Additionally, the *sicA* promoter has no activity in *E. coli* under inducing conditions, which is a useful trait for cloning purposes (Figure 3A).

The *sicA* promoter turns on within 3 h after cells are shifted into inducing media (Figure 2A) (Temme *et al*, 2008). At this point, secreted protein begins to accumulate in the media (Figure 2A). This can be clearly visualized in a Coomassie gel (Figure 3D) along with other proteins naturally secreted by *Salmonella*. There was no observable lysis or outer membrane shedding in the secretion assay, as determined using an antibody against the MalE periplasmic protein (Majander *et al*, 2005).

There are multiple chaperones that target proteins to the SPI-1 T3SS and they interact with different N-terminal tags (Table I). A secretion assay was performed for the known tag/chaperone pairs to compare the amount of protein that is exported to the supernatant. This was repeated for each protein in this study to identify the optimal combination (Figure 2B; Supplementary information S2) (Materials and methods) (Collazo and Galan, 1996). The SptP/SicP combination was found to yield the largest amount of secreted protein for DH, ADF-2 (Figure 2B), and ADF-3 (Supplementary information S2). For ADF-1, the optimum pair was SopB/SigE.

The SptP tag consists of a 15 amino-acid signal sequence and 152 amino-acid CBD (Lee and Galan, 2004). The SptP CBD interacts with SicP chaperone dimers, which direct the SptP-tagged protein to the SPI-1 T3SS (Stebbins and Galan, 2001; Akeda and Galan, 2005). It has been shown that the signal sequence directs the protein generally to T3SSs and flagella. The CBD causes the protein to be directed specifically to the SPI-1 T3SS (Lee and Galan, 2004). The T3SS secretion signal is not self-cleaving. A tobacco etch virus (TEV) protease site is added after the tag so that it can be removed after export.

Variants of the pCASP plasmid were constructed by removing the tag and chaperone parts and tested for the loss of secretion. When expressed without the N-terminal SptP tag, there is a dramatic decline in the amount of DH protein that appears in the supernatent (Figure 3A). The deletion of the tag from the silk monomers results in 100-fold less protein in the supernatant (Table II). Secretion is also reduced when the SicP chaperone is only expressed at wild-type levels and not co-transcribed on the plasmid (-SicP). We tested the effect of expressing the chaperone in conjunction with a (−) tag protein in the pCASP system. The construct was run using the secretion assay and analyzed by western blot (Figure 3B).

Secretion was also tested for *Salmonella* strains where components of the SPI-1 T3SS and flagellum have been knocked out (Figure 3C). The *sicA* promoter differs in activity for these knockouts, so a plasmid was constructed where the chaperone and tagged protein are under the control of an IPTG-inducible promoter. There was little effect when the flagella master regulators (Δ*flhDC*) are knocked out. However, when critical structural components of the SPI-1 T3SS are knocked out (Δ*prgHIJKorgABC*), secretion is eliminated. These results were confirmed using a set of earlier constructed knockout strains (Δ*invA* and Δ*flhGHI*) (Supplementary information) (Lee and Galan, 2004). Together, these data indicate that the observed protein secretion is SPI-1 dependent.

Each synthetic silk gene was ligated into the pCASP secretion plasmid with the optimal tag/chaperone pair and

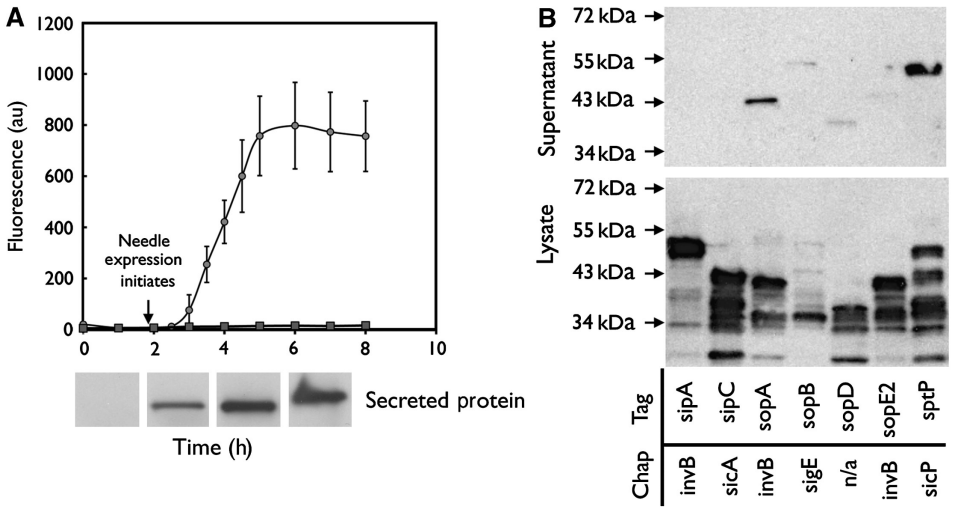


Figure 2 The dynamics of the *sicA* promoter and the testing secretion efficiency of tag–chaperone pairs. (**A**) The time-dependent activation curve of *psicA* fused to GFP in *Salmonella* SL1344 (gray circles) and *psicA* fused to GFP in *E. coli* DH10B (gray squares) as measured by flow cytometry. Error bars represent ±1 s.d. of three to five independent measurements. Cells are grown in SPI-1 repressing conditions and at $T=0$ are shifted into SPI-1 inducing media. At 2.5 h the population starts to turn on and by 3 h the majority of cells are in the on state. A western blot for secreted DH protein shows that protein in the supernatant is a function of time. An arrow has been superimposed on the graph indicating the activation of the *prgHIDKorgABC* operon (encoding the secretion needle). (**B**) The assay to find the optimal secretion tag/chaperone pair is shown. All known SPI-1 N-terminal tag and chaperone pairs are tested for the secretion of ADF-2 (Table I). Each pair was cloned into the pCASP plasmid under the control of the *sicA* promoter between the *Xho*I and *Hind*III restriction sites. A secretion assay was performed and the western blot of supernatant and lysate samples is shown. The same comparison was made for each remaining silk protein (Supplementary information).

tested for secretion (Figure 3E). Protein expression and secretion rates were determined using quantitative western blots and a Typhoon variable mode imager (Supplementary information). The secretion rates are shown in Table II. After 8 h, the total protein secreted ranges between 0.7 and 14 mg l^{-1}. After recovery, the N-terminal SptP tag can be removed by *in vitro* TEV proteolysis, leaving only two amino acids (SG) on the N-terminus of the secreted protein (Figure 3F). The total protein that was expressed ranged from 1–12 mg l^{-1} h^{-1} (Table II). The efficiency of secretion is defined as the ratio of secreted protein to expressed protein. For the silks, we observe efficiencies of 7–14%.

Discussion

The ability to export proteins directly to the extracellular environment is an important tool in cellular engineering. The secretion of silk monomers is an example of its utility for the production of recombinant proteins that form inclusion bodies or are toxic when allowed to accumulate inside of the cell. Further, there are applications in the export of enzymes that act on substrates that cannot diffuse through the cell membranes.

Here, we describe a control system that contains all of the necessary genetic parts for heterologous proteins to be secreted to the extracellular environment through the *Salmonella* SPI-1 T3SS (Box 1). At the core of this system is a feedback circuit that responds to the capacity of the bacterium to export protein. This circuit activates the transcription of an operon containing the protein to be secreted, which is fused to the SptP N-terminal secretion signal. The SicP chaperone is also encoded, which binds and directs the heterologous protein to the SPI-1 T3SS.

The yields of secreted protein using type III secretion (Table II) are as good or better than other methods of extracellular protein production (Choi and Lee, 2004). Earlier reported values range from 0.5 to 10 mg l^{-1} h^{-1} in shake flasks using a range of protein export and recovery methods (Lucic *et al*, 1998; Tong *et al*, 2000; Fu *et al*, 2003). Unlike the Sec and Tat pathways, the T3SS translocates proteins through both membranes to the extracellular environment.

Natural *Yersinia*, *Shigella*, and *Salmonella* effector proteins have been observed to accumulate up to grams per liter in the growth media. The secretion rate of the SPI-1 T3SS has been measured in individual cells to be 7–60 proteins per cell per second when secreting a natural effector (Enninga *et al*, 2005; Schlumberger *et al*, 2005). This corresponds to an approximate theoretical yield of 10 mg l^{-1} h^{-1} OD for a protein the size of ADF-2. Our system is able to achieve ~10% of this yield for the spider silk ADF-2 in a shake flask culture.

Whole gene DNA synthesis is a useful tool to make large-scale changes to genes. In this work, three spider silk genes were computationally designed for expression in eubacteria and constructed using automated DNA synthesis and assembly. Only the amino-acid sequence information is required for gene construction, rather than physical genetic material from the source organism. The degeneracy of the genetic code allows many alternative nucleotide sequences to encode the same protein. This enables synthetic genes to be designed that simultaneously match the recombinant host's codon usage, while reducing the DNA repetitiveness. As of their highly repetitive amino-acid sequences, silk monomers represent one of the most challenging targets of automated DNA synthesis and assembly. The three synthetic ADF genes built here

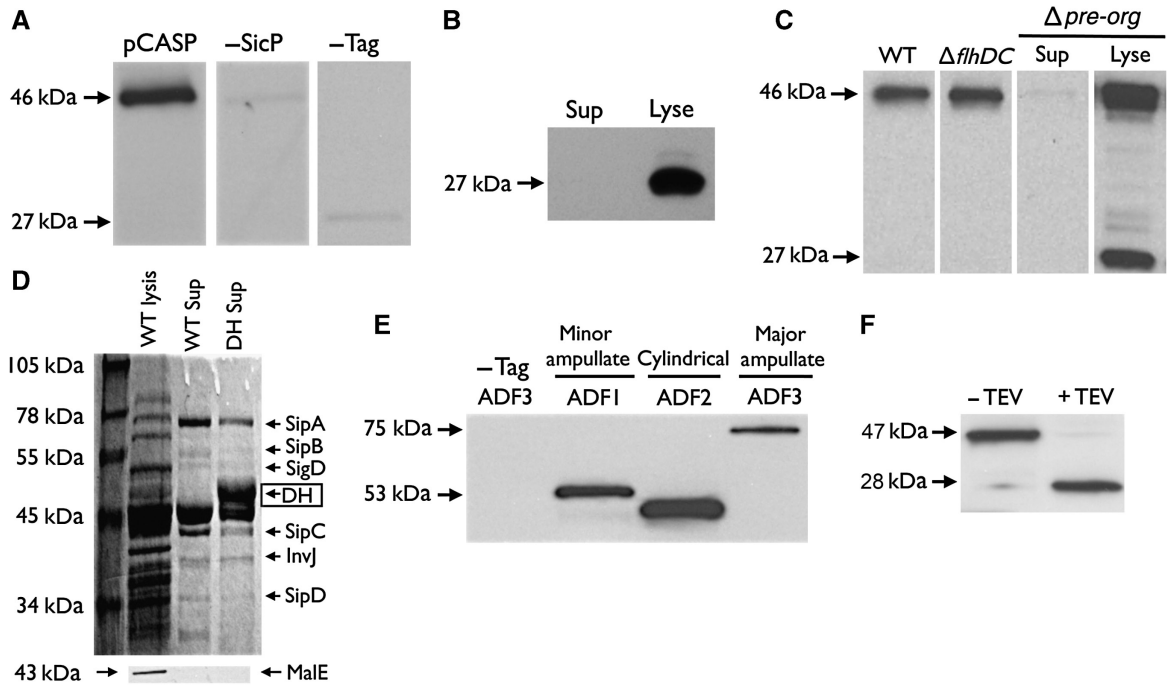


Figure 3 Secretion assays are shown for pCASP plasmid, strain variations, and silks. (**A**) The supernatant contains significant secreted DH protein for the pCASP plasmid (lane 1). Protein secretion is significantly reduced when the chaperone is not co-expressed on the plasmid (*-sicP*, lane 2) or the N-terminal SptP secretion tag is absent (−Tag, lane 3). (**B**) A western blot showing that expression of the chaperone SicP does not cause protein leakage. A variant of pCASP was generated lacking the SptP tag (+ *sicP* –*sptP*); DH protein was expressed in a secretion assay. No protein was detectable in the supernatant until it was concentrated 16 × and exposed for 1 min (data not shown). (**C**) Secretion of the DH domain from wild-type *Salmonella typhimurium* SL1344 (WT) is compared with two knockouts. There is little effect when the flagella master regulators are knocked out (Δ*flhDC*). The accumulation of protein in the supernatant (Sup) is significantly reduced when the *prg-org* operon (including *prgHIDKorgABC)* is knocked out (Δ*prg-org*), but can still be detected in the lysate (Lyse). (**D**) A Coomassie stained gel of concentrated protein (equivalent to 1.5 ml of supernatant) is shown for lysed cells (WT lysis, lane 1), culture supernatant from wild-type *Salmonella*, (WT Sup, lane 2), and cells secreting the DH domain (DH Sup, lane 3). The DH domain accumulates significantly in the supernatant. The identity of the DH band was confirmed by immunoprecipitation of the protein from supernatant (not shown). The pattern of other secreted proteins in the supernatant matches those reported earlier (Collazo and Galan, 1996). Below the Coomassie gel, a western blot is shown. 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. (**E**) A secretion assay is shown for the synthetic ADF-1, -2, and -3 genes (lanes 2–5). When the N-terminal secretion tag is removed from the sequence (−Tag), no protein is detected in the supernatant (lane 1). The secretion yields with and without the N-terminal tag are determined using quantitative westerns (Supplementary information) and presented in Table II. (**F**) TEV protease cleaves the SptP secretion tag from the silk proteins. ADF-2 before digestion (lane 1) is reduced in size by 19 kDa when the SptP secretion tag is removed by TEV protease (lane 2).

Table I SPI-1 N-terminal tags and chaperones

Tag	Length[a]	Chaperone	Part numbers[b]	References
SipA	169	InvB	BBa_J64035	Bronstein *et al* (2000)
SipC	167	SicA	BBa_J64040	Tucker and Galan (2000)
SopA	96	InvB	BBa_J64037	Higashide and Zhou (2006)
SopE2	105	InvB	BBa_J64038	Ehrbar *et al* (2003) and Karavolos *et al* (2005)
SptP	167	SicP	BBa_J64008	Fu and Galan (1998)
SopB	168	SigE	BBa_J64041	Hong and Miller (1998) and Knodler *et al* (2006)
SopD	40	none	BBa_J64042	Wood *et al* (2004) and Zhang *et al* (2002)

[a]Number of N-terminal amino acids included in the tag.
[b]The part number from the Registry of Standard Biological Parts (http://www.partsregistry.org/). These numbers correspond to superparts that include the chaperone, ribosome-binding site, and the N-terminal tag.

represent all of the known sequence information for these silks, but they are still fragments of the complete silk genes.

The ability to construct the wild-type genes from information alone will make it possible to further explore—and modify—the amazing diversity of natural materials. More broadly, declining automated synthesis costs make it possible to rapidly construct large libraries of proteins, enzymes, or pathways from many diverse organisms using only information retrieved from sequence databases (Bayer *et al*, 2009). This use of synthetic metagenomics could be particularly applicable to areas of material space such as spider silks, when cDNA is difficult to obtain from the natural source and when isolated is unstable in recombinant hosts. Concurrently, it enables the large-scale modification of the sequences for expression in a recombinant host. This approach will revolutionize how natural diversity is explored when engineering cells.

Table II Yields of expressed and secreted silk (nmol l^{-1} h^{-1})[a]

		+Tag[b]		−Tag	
	Mass (kDa)	Super[c]	Lysate	Super[d]	Lysate
ADF-1	30.8	2.9 ± 1.0	31 ± 5	0.01 ± 0.006	34 ± 5
ADF-2	25.0	70 ± 15	410 ± 80	0.4 ± 0.1	200 ± 30
ADF-3	56.2	6.9 ± 0.4	90 ± 21	0.02 ± 0.006	68 ± 22

[a]The reported numbers are the mean of five (four, −tag ADF-2) independent measurements and the error is 1 s.d.
[b]Rates are shown when the silk monomer is fused to the N-terminal secretion tag (+tag) and when expressed without the tag (−tag).
[c]Rates for the secreted protein collected in the supernatant (super) and nonsecreted protein (lysate) are shown.
[d]When expressed without the tag, the protein is only detectible in the supernatant upon 30 × concentration (Materials and methods).

Materials and methods

See Supplementary information for detailed Materials and methods.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Jorge Galan (Yale), George Georgiou (UT-Austin), and Wendell Lim (UCSF) for providing genetic material and strains. DMW was supported by a Genentech-Sandler Graduate Fellowship. EAM was supported by an NSF Graduate Research Fellowship and a National Defence Science and Engineering Graduate Fellowship administered by the American Society for Engineering Education. CAV was supported by a Pew Fellowship, the Office of Naval Research, Packard Fellowship, NIH EY016546, NIH AI067699, NSF BES-0547637, UC-Discovery, a Sandler Family Opportunity Award, and the SynBERC NSF ERC (www.synberc.org).

Conflict of interest

The authors declare that they have no conflict of interest.

References

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

Altier C (2005) Genetic and environmental control of Salmonella invasion. *J Microbiol* **43:** 85–92

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

Bajaj V, Lucas RL, Hwang C, Lee CA (1996) Co-ordinate regulation of Salmonella typhimurium invasion genes by environmental and regulatory factors is mediated by control of hilA expression. *Mol Microbiol* **22:** 703–714

Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA (2009) Synthesis of methyl halides from biomass using engineered microbes. *J Am Chem Soc* **131:** 6508–6515

Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S (2005) Prediction of twin-arginine signal peptides. *BMC Bioinformatics* **6:** 167

Berks BC (1996) A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol* **22:** 393–404

Boyd AP, Grosdent N, Totemeyer S, Geuijen C, Bleves S, Iriartes M, Lambermont I, Octave JN, Cornelis GR (2000) Yersinia enterocolitica can deliver Yop proteins into a wide range of cell types: Development of a delivery system for heterologous proteins. *Eur J Cell Biol* **79:** 659–671

Bronstein PA, Miao EA, Miller SI (2000) InvB is a type III secretion chaperone specific for SspA. *J Bacteriol* **182:** 6638–6644

Chen LM, Briones G, Donis RO, Galan JE (2006) Optimization of the delivery of heterologous proteins by the Salmonella enterica serovar Typhimurium type III secretion system for vaccine development. *Infect Immun* **74:** 5826–5833

Choi JH, Lee SY (2004) Secretory and extracellular production of recombinant proteins using Escherichia coli. *Appl Microbiol Biotechnol* **64:** 625–635

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

Darwin KH, Miller VL (1999) InvF is required for expression of genes encoding proteins secreted by the SPI1 type III secretion apparatus in Salmonella typhimurium. *J Bacteriol* **181:** 4949–4954

Darwin KH, Miller VL (2000) The putative invasion protein chaperone SicA acts together with InvF to activate the expression of Salmonella typhimurium virulence genes. *Mol Microbiol* **35:** 949–960

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

DeLisa MP, Tullman D, Georgiou G (2003) Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc Natl Acad Sci USA* **100:** 6115–6120

Economou A, Wickner W (1994) Seca promotes preprotein translocation by undergoing Atp-driven cycles of membrane insertion and deinsertion. *Cell* **78:** 835–843

Ehrbar K, Friebel A, Miller SI, Hardt WD (2003) Role of the Salmonella pathogenicity island 1 (SPI-1) protein InvB in type III secretion of SopE and SopE2, two Salmonella effector proteins encoded outside of SPI-1. *J Bacteriol* **185:** 6950–6967

Eichelberg K, Galan JE (1999) Differential regulation of Salmonella typhimurium type III secreted proteins by pathogenicity island 1 (SPI-1)-encoded transcriptional activators InvF and hilA. *Infect Immun* **67:** 4099–4105

Ellermeier CD, Ellermeier JR, Slauch JM (2005) HilD, HilC and RtsA constitute a feed forward loop that controls expression of the SPI1 type three secretion system regulator hilA in Salmonella enterica serovar Typhimurium. *Mol Microbiol* **57:** 691–705

Enninga J, Mounier J, Sansonetti P, Tran Van Nhieu G (2005) Secretion of type III effectors into host cells in real time. *Nat Methods* **2:** 959–965

Fahnestock SR, Irwin SL (1997) Synthetic spider dragline silk proteins and their production in Escherichia coli. *Appl Microbiol Biotechnol* **47:** 23–32

Feldman MF, Muller S, Wuest E, Cornelis GR (2002) SycE allows secretion of YopE-DHFR hybrids by the Yersinia enterocolitica type III Ysc system. *Mol Microbiol* **46:** 1183–1197

Fu Y, 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

Fu Z, Hamid SB, Razak CN, Basri M, Salleh AB, Rahman RN (2003) Secretory expression in Escherichia coli and single-step purification of a heat-stable alkaline protease. *Protein Expr Purif* **28:** 63–68

Galan JE, Collmer A (1999) Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284:** 1322–1328

Georgiou G, Segatori L (2005) Preparative expression of secreted proteins in bacteria: status report and future prospects. *Curr Opin Biotechnol* **16:** 538–545

Gerchman Y, Weiss R (2004) Teaching bacteria a new language. *Proc Natl Acad Sci USA* **101:** 2221–2222

Gosline JM, Guerette PA, Ortlepp CS, Savage KN (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function. *J Exp Biol* **202:** 3295–3303

Guerette PA, Ginzinger DG, Weber BH, Gosline JM (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science* **272:** 112–115

Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. *Trends Biotechnol* **22:** 346–353

Harvey BR, Georgiou G, Hayhurst A, Jeong KJ, Iverson BL, Rogers GK (2004) Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries. *Proc Natl Acad Sci USA* **101:** 9193–9198

Higashide W, Zhou D (2006) The first 45 amino acids of SopA are necessary for InvB binding and SPI-1 secretion. *J Bacteriol* **188:** 2411–2420

Hofmann S, Hagenmuller H, Koch AM, Muller R, Vunjak-Novakovic G, Kaplan DL, Merkle HP, Meinel L (2007) Control of *in vitro* tissue-engineered bone-like structures using human mesenchymal stem cells and porous silk scaffolds. *Biomaterials* **28:** 1152–1162

Hong KH, Miller VL (1998) Identification of a novel Salmonella invasion locus homologous to Shigella ipgDE. *J Bacteriol* **180:** 1793–1802

Huemmerich D, Helsen CW, Quedzuweit S, Oschmann J, Rudolph R, Scheibel T (2004a) Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* **43:** 13604–13612

Huemmerich D, Scheibel T, Vollrath F, Cohen S, Gat U, Ittah S (2004b) Novel assembly properties of recombinant spider dragline silk proteins. *Curr Biol* **14:** 2070–2074

Hussain NK, Jenna S, Glogauer M, Quinn CC, Wasiak S, Guipponi M, Antonarakis SE, Kay BK, Stossel TP, Lamarche-Vane N, McPherson PS (2001) Endocytic protein intersectin-l regulates actin assembly via Cdc42 and N-WASP. *Nat Cell Biol* **3:** 927–932

Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* **292:** 2080–2083

Karavolos MH, Roe AJ, Wilson M, Henderson J, Lee JJ, Gally DL, Khan CM (2005) Type III secretion of the Salmonella effector protein SopE is mediated via an N-terminal amino acid signal and not an mRNA sequence. *J Bacteriol* **187:** 1559–1567

Kluge JA, Rabotyagova U, Leisk GG, Kaplan DL (2008) Spider silks and their applications. *Trends Biotechnol* **26:** 244–251

Knodler LA, Bertero M, Yip C, Strynadka NC, Steele-Mortimer O (2006) Structure-based mutagenesis of SigE verifies the importance of hydrophobic and electrostatic residues in type III chaperone function. *Mol Microbiol* **62:** 928–940

Konjufca V, Wanda SY, Jenkins MC, Curtiss III R (2006) A recombinant attenuated Salmonella enterica serovar Typhimurium vaccine encoding Eimeria acervulina antigen offers protection against E acervulina challenge. *Infect Immun* **74:** 6785–6796

Kubori T, Galan JE (2002) Salmonella type III secretion-associated protein InvE controls translocation of effector proteins into host cells. *J Bacteriol* **184:** 4699–4708

Kubori T, Matsushima Y, Nakamura D, Uralil J, Lara-Tejero M, Sukhan A, Galan JE, Aizawa SI (1998) Supramolecular structure of the Salmonella typhimurium type III protein secretion system. *Science* **280:** 602–605

Lazaris A, Arcidiacono S, Huang Y, Zhou JF, Duguay F, Chretien N, Welsh EA, Soares JW, Karatzas CN (2002) Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* **295:** 472–476

Lee PA, Tullman-Ercek D, Georgiou G (2006) The bacterial twin-arginine translocation pathway. *Annu Rev Microbiol* **60:** 373–395

Lee SH, Galan JE (2004) Salmonella type III secretion-associated chaperones confer secretion-pathway specificity. *Mol Microbiol* **51:** 483–495

Lee SM, Pippel E, Gosele U, Dresbach C, Qin Y, Chandran CV, Brauniger T, Hause G, Knez M (2009) Greatly increased toughness of infiltrated spider silk. *Science* **324:** 488–492

Lee VT, Schneewind O (2002) Yop fusions to tightly folded protein domains and their effects on Yersinia enterocolitica type III secretion. *J Bacteriol* **184:** 3740–3745

Lewis RV (2006) Spider silk: ancient ideas for new biomaterials. *Chem Rev* **106:** 3762–3774

Lucas RL, Lee CA (2000) Unravelling the mysteries of virulence gene regulation in Salmonella typhimurium. *Mol Microbiol* **36:** 1024–1033

Lucic MR, Forbes BE, Grosvenor SE, Carr JM, Wallace JC, Forsberg G (1998) Secretion in Escherichia coli and phage-display of recombinant insulin-like growth factor binding protein-2. *J Biotechnol* **61:** 95–108

Lundberg U, Vinatzer U, Berdnik D, von Gabain A, Baccarini M (1999) Growth phase-regulated induction of Salmonella-induced macrophage apoptosis correlates with transient expression of SPI-1 genes. *J Bacteriol* **181:** 3433–3437

Majander K, Anton L, Antikainen J, Lang H, Brummer M, Korhonen TK, Westerlund-Wikstrom B (2005) Extracellular secretion of polypeptides using a modified Escherichia coli flagellar secretion apparatus. *Nat Biotechnol* **23:** 475–481

Marlovits TC, Kubori T, Lara-Tejero M, Thomas D, Unger VM, Galan JE (2006) Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* **441:** 637–640

Marlovits TC, Kubori T, Sukhan A, Thomas DR, Galan JE, Unger VM (2004) Structural insights into the assembly of the type III secretion needle complex. *Science* **306:** 1040–1042

Nielsen H, Engelbrecht J, Brunak S, vonHeijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10:** 1–6

Pohlschroder M, Hartmann E, Hand NJ, Dilks K, Haddad A (2005) Diversity and evolution of protein translocation. *Annu Rev Microbiol* **59:** 91–111

Prince JT, McGrath KP, DiGirolamo CM, Kaplan DL (1995) Construction, cloning, and expression of synthetic genes encoding spider dragline silk. *Biochemistry* **34:** 10879–10885

Rising A, Nimmervoll H, Grip S, Fernandez-Arias A, Storckenfeldt E, Knight DP, Vollrath F, Engstrom W (2005) Spider silk proteins—mechanical property and gene sequence. *Zoolog Sci* **22:** 273–281

Rodrigue A, Chanal A, Beck K, Muller M, Wu LF (1999) Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway. *J Biol Chem* **274:** 13223–13228

Russmann H, Kubori T, Sauer J, Galan JE (2002) Molecular and functional analysis of the type III secretion signal of the Salmonella enterica InvJ protein. *Mol Microbiol* **46:** 769–779

Russmann H, Shams H, Poblete F, Fu Y, Galan JE, Donis RO (1998) Delivery of epitopes by the Salmonella type III secretion system for vaccine development. *Science* **281:** 565–568

Sargent F, Bogsch EG, Stanley NR, Wexler M, Robinson C, Berks BC, Palmer T (1998) Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO J* **17:** 3640–3650

Schlumberger MC, Muller AJ, Ehrbar K, Winnen B, Duss I, Stecher B, Hardt WD (2005) Real-time imaging of type III secretion: Salmonella SipA injection into host cells. *Proc Natl Acad Sci USA* **102:** 12548–12553

Seidel A, Liivak O, Calve S, Adaska J, Ji G, Yang Z, Grubb D, Zax D, Jelinski L (2000) Regenerated spider silk: processing, properties and structure. *Macromolecules* **33:** 775–780

Somerville C (2007) Biofuels. *Curr Biol* **17:** R115–R119

Sorensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J Biotechnol* **115:** 113–128

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

Swanson BO, Blackledge TA, Summers AP, Hayashi CY (2006) Spider dragline silk: correlated and mosaic evolution in high-performance biological materials. *Evolution* **60:** 2539–2551

Tartera C, Metcalf ES (1993) Osmolarity and growth-phase overlap in regulation of Salmonella typhi adherence to and invasion of human intestinal cells. *Infect Immun* **61:** 3084–3089

Temme K, Salis H, Tullman-Ercek D, Levskaya A, Hong SH, Voigt CA (2008) Induction and relaxation dynamics of the regulatory network controlling the type III secretion system encoded within Salmonella pathogenicity island 1. *J Mol Biol* **377:** 47–61

Tong L, Lin Q, Wong WK, Ali A, Lim D, Sung WL, Hew CL, Yang DS (2000) Extracellular expression, purification, and characterization of a winter flounder antifreeze polypeptide from Escherichia coli. *Protein Expr Purif* **18:** 175–181

Tucker SC, Galan JE (2000) Complex function for SicA, a Salmonella enterica serovar typhimurium type III secretion-associated chaperone. *J Bacteriol* **182:** 2262–2268

Vegh BM, Gal P, Dobo J, Zavodszky P, Vonderviszt F (2006) Localization of the flagellum-specific secretion signal in Salmonella flagellin. *Biochem Biophys Res Commun* **345:** 93–98

Vendrely C, Scheibel T (2007) Biotechnological production of spider-silk proteins enables new applications. *Macromol Biosci* **7:** 401–409

Vollrath F, Knight DP (2001) Liquid crystalline spinning of spider silk. *Nature* **410:** 541–548

Wang Y, Blasioli DJ, Kim HJ, Kim HS, Kaplan DL (2006a) Cartilage tissue engineering with silk scaffolds and human articular chondrocytes. *Biomaterials* **27:** 4434–4442

Wang Y, Kim HJ, Vunjak-Novakovic G, Kaplan DL (2006b) Stem cell-based tissue engineering with silk biomaterials. *Biomaterials* **27:** 6064–6082

Wickner W, Schekman R (2005) Protein translocation across biological membranes. *Science* **310:** 1452–1456

Wittrup KD (2000) The single cell as a microplate well. *Nat Biotechnol* **18:** 1039–1040

Wood MW, Williams C, Upadhyay A, Gill AC, Philippe DL, Galyov EE, van den Elsen JM, Bagby S (2004) Structural analysis of Salmonella enterica effector protein SopD. *Biochim Biophys Acta* **1698:** 219–226

Zhang S, Santos RL, Tsolis RM, Stender S, Hardt WD, Baumler AJ, Adams LG (2002) The Salmonella enterica serotype typhimurium effector proteins SipA, SopA, SopB, SopD, and SopE2 act in concert to induce diarrhea in calves. *Infect Immun* **70:** 3843–3855,

Zhang YHP, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv* **24:** 452–481

Zhou SD, Yomano LP, Saleh AZ, Davis FC, Aldrich HC, Ingram LO (1999) Enhancement of expression and apparent secretion of Erwinia chrysanthemi endoglucanase (encoded by celZ) in Escherichia coli B. *Appl Environ Microbiol* **65:** 2439–2445