Title
Engineering the Salmonella Type III Secretion System

Author
Widmaier, Dan

Publication Date
2010

Peer reviewed|Thesis/dissertation

Engineering the *Salmonella* Type III Secretion System

by

Daniel Matthew Widmaier

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements

The text of Chapter 2 of this dissertation is a revised reprint of the material as it appears in the journal Molecular Systems Biology. The coauthor Christopher Voigt listed in this publication directed and supervised the research that forms the basis for the dissertation. Coauthors Danielle Tullman-Ercek, Ethan Mirsky, and Rena Hill assisted in designing and running aspects of the experiments. Coauthors Sridhar Govindarajan and Jeremy Minshull performed the silk gene sequence optimization and synthesis.

First thanks go out to my thesis advisor Chris Voigt for his guidance, advice, and project management as this dissertation went from impossible idea to published story.

Thank-you to my thesis committee, Wendell Lim and Jim Wells for their guidance through our discussions and thesis committee meeting.

I have fond memories of Chris Anderson and Anselm Levskaya for teaching me to clone in the early days of the Voigt lab.

Thank you for the rest of my CCB class, the program, and the administrators (Chris Olson in particular) for a memorable trip through graduate school

A huge thanks go out to Anselm, Jeff, Travis, Liz, Eli, Karsten, Alvin, Danielle, Ethan, Anderson, Soon-Ho, Felix, Howard, George, Rena, Dehua and everyone else in the voigt lab for making it a vibrant and creative place to work. I will always remember the interesting adventures we've had in and out of lab and friendships created.

A special acknowledgement to David and Ethan for our future in business at Refactored Materials that was born out of random encounters from graduate school.

Thanks to my family for their support and understanding while I continued my education for another 6 years.

Otto for putting up with me being gone all the time, but brightening my day when I come home.

Finally, an extra special thank-you to my girlfriend Aimee and for being so understanding throughout.

Abstract

Engineering the *Salmonella* Type III Secretion System

Daniel M. Widmaier

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 (SPI-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 an array of spider silk monomers 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} $hour^{-1}$ 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.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1. Engineering Living Systems

The focus of this dissertation is to examine a complex molecular machine, the Type III Secretion System (T3SS) located at *Salmonella* Pathogeneity Island 1 (SPI-1), and reprogram its function for the secretion of reocombinant proteins. SPI-1 is a complex genetic machine consisting of over 30 genes clustered at a single 34 kilobase locus of DNA. SPI-1 encodes a range of functional products ranging from transcription factors, signaling proteins, structural proteins to drive the secretion complex, secreted effector proteins and molecular chaperones. The complexity of this system combined with interesting assembly dynamics make it an idea target for engineered protein secretion.

Protein secretion is a difficult and interesting problem in many biotechnology applications. The ability to direct secretion of proteins outside the cell simplifies the recovery and purification, if protein is the desired product. Secretion also enables delivering enzymes to substrates that cannot pass through the cellular membranes such as cellulosic biomass. In this dissertation the ability to connect secretion capacity to

heterologous protein expression is demonstrated and as a first iteration system shows promise for future improvement.

The second focus of this dissertation is to demonstrate how external cellular logic is connected to SPI-1. Natural transcriptional logic in the *psicA* promoter of SPI-1 is expanded by adding sensitivity to the lac control system. This addition of control does not disturb the natural ability of *psicA* to function under many circumstances. However the presence of repressor protein (LacI) and absence of small molecule inducer (IPTG) turn off the expression of secreted protein strongly. This ability to add additional control to the system makes it possible to reconfigure the secretion of protein in response to specific cellular inputs, namely the presence of a functional secretion system and synthetic genetic circuits that control the induction or degradation of repressor protein.

The protein of interest for secretion throughout this dissertation is the spider silk family of proteins. Spider's silk is a protein material capable of making remarkable fibers with mechanical properties exceeding that of the most sophisticated man made polymers. The production of synthetic spider silk is a landmark challenge in the field of biomimetic research and remains an open problem. One key limitation in this field is the availability of recombinant protein to test hypotheses about fiber formation and composition. The intent of this dissertation is to begin to shed some light on new innovative approaches to solving this problem.

1.2. Outline of this dissertation

This dissertation is broken down into two primary chapters:

Chapter 2 is a study of the function of SPI-1 and how to utilize natural signal processing to drive the secretion of spider silk protein.

Chapter 3 is a study in how to determine the secretion system constraints for spider silk production and add external control to transcriptional logic in SPI-1.

Chapter 2
Understanding and Reprogramming Type III Secretion

2.1. 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. 2004).

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 upon 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 appears 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 *E. coli* and used to secrete *Erwinia* cellulases (Zhou et al. 1999). Alternatively, gram-positive bacteria offer a single secretion event to the extra cellular space and offer an attractive platform for secretion engineering.

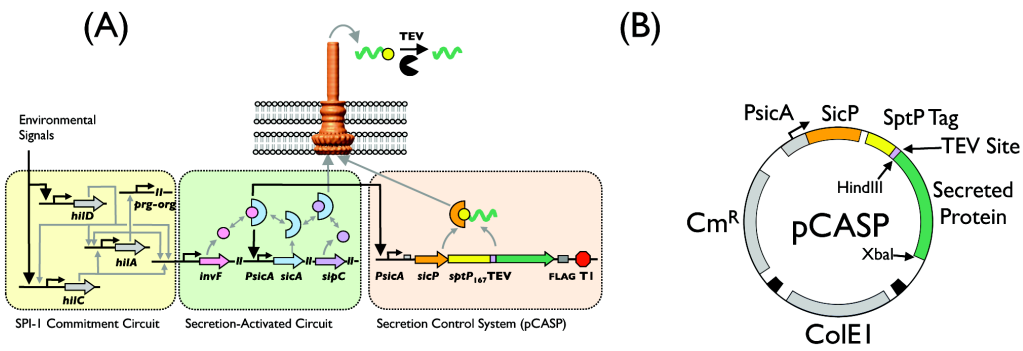


Figure 1

The overall goal is to design a control system that interfaces with the SPI-1 T3SS to export heterologous proteins at the proper time during the growth cycle. 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 drives 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 **(panel 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).

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 prior to 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 (Figure 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; Darwin and Miller 2000; Darwin and Miller 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; Darwin and Miller 2000; Tucker and Galan 2000; Darwin and Miller 2001).

The *Salmonella* SPI-1 T3SS has several properties that make it a good tool for the secretion of proteins of biotechnological interest. It is highly expressed under standard laboratory conditions (Luria-Bertani Broth at 37^{o}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, 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). Based on this previous work, we have constructed a system that contains all of the necessary genetic parts to secrete heterologous proteins (Figure 1A). 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. 2006; Wang et al. 2006; 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 due to 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). Fourth, 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. 2004). Finally, due to the repetitive nature of the silk genes and the long length (~10kB) the sequencing of wild-type silk genes is limited almost entirely to partial fragments of genes. Here we demonstrate that the SPI-1 T3SS can be harnessed to express and export silk proteins from the cell into the extracellular environment.

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. 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.

2.2. Results

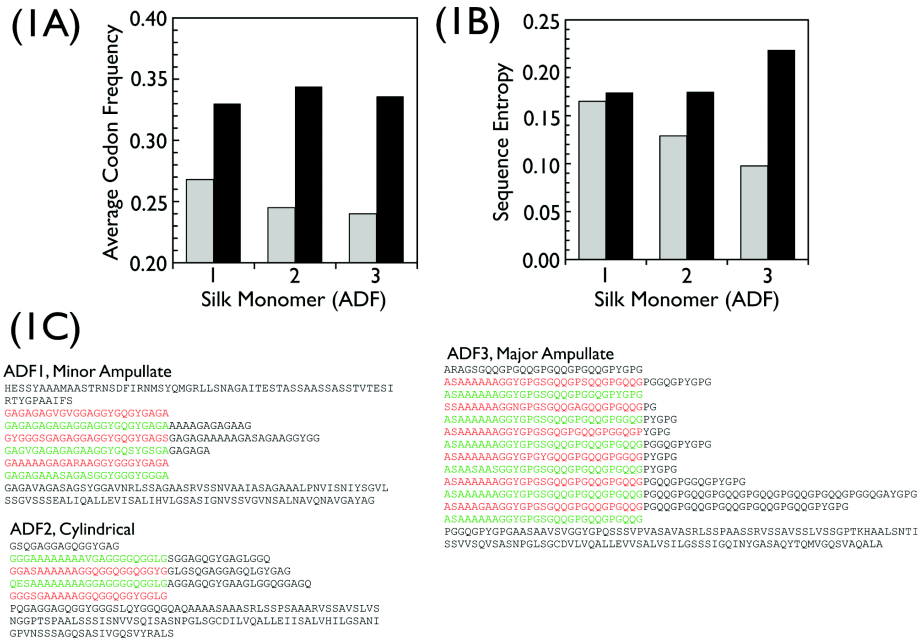


Figure 2

Changes made to the *Araneus* spider silk coding sequences during synthesis. **(A)** Codon frequency is the frequency with which a given codon triplet is used relative to all possible triplets for that amino acid. The average frequency is the mean of the codon frequencies across the entire gene sequence. The average codon frequency is shown for the spider (grey) 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 (grey) and synthetic (black) genes. The repeat units for each silk monomer were manually aligned (Supplementary Information) and the sequence entropy is

calculated: $S = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{4} p_i(j)\ln p_i(j)$, 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 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 it is produced in. The repetitive regions used in the sequence entropy calculations are in red and green. The sequences shown match all known sequence information about the three silks.

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 2**C). 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 2**A, 2B). 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 (Figure 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; Darwin and Miller 2000; Darwin and Miller 2001; Temme et al. 2008). The output of this circuit (*PsicA*) drives the expression of a chaperone and an N-terminally tagged protein to be secreted. The *sicA* promoter drives the transcription of an operon containing the heterologous protein fused to an N-terminal secretion tag and the associated chaperone (Figure 1). The human DH domain (Hussain et al. 2001) was chosen as an initial heterologous protein because it is small (24kD), 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 (Figure 1) and are available in the Registry of Standard Biological Parts (http://www.partsregistry.org/). The part numbers (BBa_) are provided throughout the text.

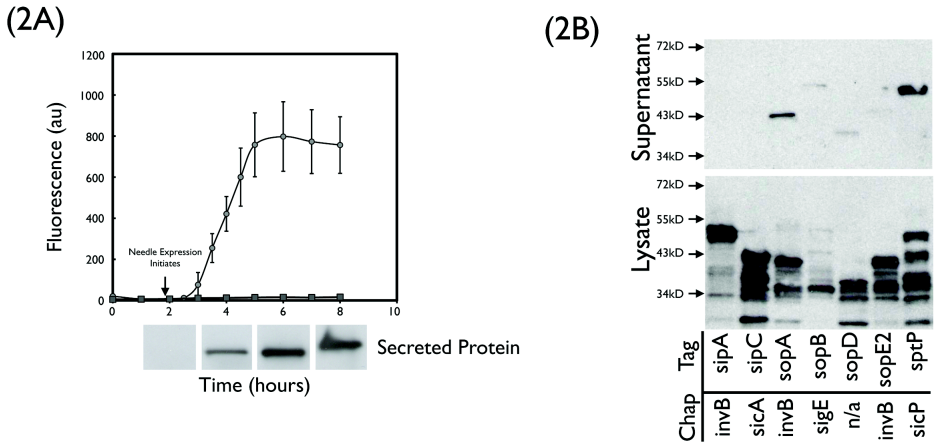


Figure 3

The function of the *sicA* promoter and testing secretion efficiency. The *sicA* promoter is one of the last promoters to be activated in the SPI-1 pathway(Temme et al. 2008). The activation time for this promoter was determined by transcriptionally fusing it to green fluorescent protein and measuring the fluorescence

using flow cytometry (Temme et al. 2008). **(A)** The time dependent activation curve of *psicA* fused to GFP in *Salmonella* SL1344 (grey circles) and *psicA* fused to GFP in *E. coli* DH10B (grey squares) as measured by flow cytometry. Error bars represent ±1 SD of 3-5 independent measurements. The *E.coli* error bars are occluded by the point markers. Cells are grown in SPI-1 repressing conditions and at T=0 are shifted into SPI-1 Inducing Media (Supplementary Information, Materials and Methods). At 2.5 hours the population starts to turn on and by 3 hours the majority of cells are in the on state and remain on until the experimental endpoint at 8 hours. A western blot for secreted DH protein shows that protein in the supernatant matches the promoter function. An arrow has been super-imposed on the graph indicating the activation of the *prgHIDKorgABC* operon (encoding the secretion needle). The activation time of the *prg* operon was adapted from previously published data(Temme et al. 2008). **(B)** Assaying to find the optimal secretion tag / chaperone pair for maximum protein throughput. All known SPI-1 N-terminal tag and chaperone pairs are tested for secretion of ADF2 silk (Table 1). Each pair was cloned into the pCASP plasmid under the control of the *sicA* promoter between the XhoI and HindIII restriction sites. A secretion assay was performed and the Western blot of supernatant and lysate samples is shown (Supplementary Information, Materials and Methods). The same comparison was made for each remaining silk protein (Supplementary Information).

The SicA:InvF gene circuit forms the core of our genetic system to secrete heterologous proteins (Figure 1). 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; Darwin and Miller 2000; Darwin and Miller 2001). This links expression to secretion capacity and can reduce the build-up of protein prior to secretion. The second advantage is that no exogenous inducer is required, the culture automatically recognizes when conditions are appropriate for protein expression and turns on automatically. The circuit can be maintained in the off state when cells are grown in low salt LB (L Broth) and then turned on after the cells are induced by high salt LB (Inducing

Medium) and the needle is constructed (Figure 3A) (Materials and Methods). This shift in the media osmolarity has been previously shown to induce the expression of SPI-1 (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). When fully active, the strength of the *sicA* promoter is equivalent to the commonly used P_{TRC} promoter on a similar plasmid with 0.01mM IPTG (Supplementary Information).

The *sicA* promoter turns on within three hours after cells are shifted into media that induces the expression of the SPI-1 T3SS (0.3M NaCl LB broth) (Figure 3A)(Temme et al. 2008). At this point, secreted protein begins to accumulate in the media (Figure 3A). This can be clearly visualized in a Coomassie gel (Figure 4D) 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).

Table 1 N-Terminal Secretion Tags Used

Tag	Length[a]	Chaperone	Part Numbers[b]	Reference
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; Karavolos et al. 2005)*
SptP	167	SicP	BBa_J64008	*(Fu and Galan 1998)*
SopB	168	SigE	BBa_J64041	*(Hong and Miller 1998; Knodler et al. 2006)*
SopD	40	none	BBa_J64042	*(Zhang et al. 2002; Wood et al. 2004)*

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.

There are multiple chaperones that target proteins to the SPI-1 T3SS and they interact with different N-terminal tags (Table 1). 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 heterologous protein in this study to identify the optimal combination (Figure 3B, Figure 6) (Materials and Methods) (Collazo and Galan 1996). The SptP/SicP combination was found to yield the largest amount of secreted protein for DH, ADF2 (Figure 3B), and ADF3 (Figure 6). For ADF1, the optimum pair was SopB/SigE (Figure 6).

The SptP tag consists of a 15 amino acid signal sequence and 152 amino acid chaperone binding domain (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 flagellum. 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, so a tobacco etch virus (TEV) protease site is added after the tag so that it can be removed after export. The *sicA* promoter, *sicA* ribosome binding site, SicP chaperone, SptP ribosome binding site and N-terminal tag, and TEV protease site were combined to construct the pCASP system (BBa_J64032) for heterologous protein secretion (Figure 1).

Table 2 Yields of Expressed and Secreted Silk (nanomoles/L-hr)[a]

		+ tag[b]		- tag	
	mass (kD)	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[‡] independent measurements and the error is one standard deviation. [‡]No Tag ADF2 Lysate consists of 4 independent measurements
b. Rates are shown when the silk monomer is fused to the N-terminal secretion tag (+tag) and when expressed without the tag (-tag).
b. Rates for the secreted protein collected in the supernatant (super) and non-secreted protein (lysate) are shown.
c. When expressed without the tag, the protein is only detectible in the supernatant upon 30x concentration (Materials and Methods).

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 4A). The deletion of the tag from the silk monomers results in 100-fold less protein in the supernatant (Table 2). 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 (Materials and Methods) and analyzed by western blot (Figure 4B).

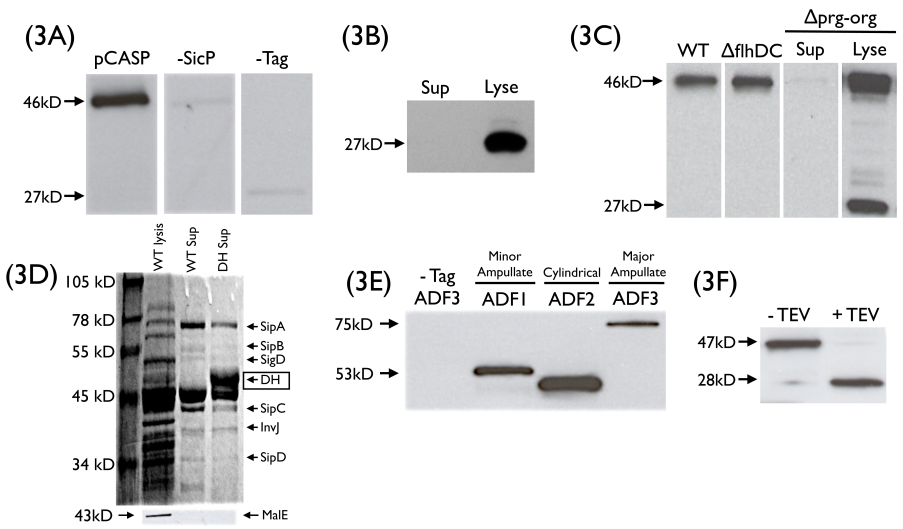


Figure 4

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 16x and exposed for 1 minute (Data not shown). **(C)** Secretion of the DH domain from wild-type *Salmonella typhimurium* SL1344 (WT) is compared to 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 previously reported (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 2. **(F)** TEV protease cleaves the SptP secretion tag from the silk proteins. ADF-2 prior to digestion (lane 1) is reduced in size by 19kD when the SptP secretion tag is removed by TEV protease (lane 2).

Secretion was also tested for *Salmonella* strains where components of the SPI-1 T3SS and flagellum have been knocked out (Figure 4C). 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 (Materials and Methods). 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 previously constructed knockout strains (Δ*invA* and Δ*flhGHI*) (Supplementary Information) (Lee and Galan 2004). Together, this data indicates 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 tested for secretion (Figure 4E) (Materials and Methods). 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 2. After 8 hours, the total protein secreted ranges between 0.7-14 mg/L, depending on the intrinsic expression rate. 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 4F). The total protein that was expressed ranged from 1 – 12 mg/L-hr (Table 2). This is consistent with the amount of protein that was recovered from the expression of consensus repeat sequences (Vendrely and Scheibel 2007). The efficiency of secretion is defined as the ratio of secreted protein to expressed protein. For the silks, we observe efficiencies of 7-14%.

2.3. Materials and Methods

2.3.1. Strains and Knockouts.

Salmonella typhimurium SL1344 (gift of Stanley Falkow, Stanford) was used for all secretion experiments and as the parent for genomic knockouts. The numbering for the deleted genomic regions corresponds to GenBank NC_003198. A flagella knockout was created by deleting the FlhDC transcriptional activators (2022067-2021136) (Δ*flhDC*::KanR). A SPI-1 T3SS knockout was constructed by deleting the *prg-org* operon (*prgHIJKorgABC*) (2869534-3013401), including the structural components of the inner membrane ring, to create strain Δ*prg-org*::KanR. Both knockouts were constructed using the method of Datsenko and Wanner (Datsenko and Wanner 2000).

2.3.2. Growth Media.

Three medias were used for cultures: Luria-Bertani Miller (LB) (Difco Cat#: 244510) Luria-Bertani Lennox (L) (Difco Cat#:241210) and SPI-1 Inducing Media (IM) (LB, 0.3M NaCl (Tartera and Metcalf 1993)). All plasmid manipulations were done in Luria-Bertani media. Media was supplemented with 25 μg/mL Kanamycin, 30 μg/mL Choloramphenicol or 100 μg/mL Ampicillin as needed.

2.3.3. Construction of pCASP, +sicP/-tag and –sicP/-tag Derivatives.

The pCASP plasmid (Genbank ID: EF179157) was constructed based on pPROTet.133 backbone (Cm^R, ColE1) (BD Clonetech) (Lutz and Bujard 1997). Part numbers are provided for the Registry of Standard Biological Parts (www.partsregistry.org). The *sicA* promoter (defined as 165 bases upstream of the *sicA* start codon) (BBa_J64001), *sicP* gene (BBa_J64002) and first 167 amino acids of *sptP* (BBa_J64003), were amplified from *S. typimuirum* genomic DNA. After being fused to the *sicA* promoter, the *sicP* gene is under the control of the *sicA* ribosome binding site, including a TTG→ATG mutation. The TEV protease cleavage sequence (GENLYFQSG, flanked by glycines for flexibility, BBa_J64007) was inserted between the SptP tag and the HindIII restriction site. The *PsicA-sicP-sptP* fragment was PCR amplified with primers including the TEV recognition sequence and the insert was re-ligated into the XhoI and HindIII restriction sites. (Kapust et al. 2002). A FLAG epitope tag (DYKDDDDK, BBa_J64005) was introduced at the XbaI site by ordering complementary oligonuclotides encoding a 3x repeat FLAG epitope tag, annealing the oligos and non-directional insertion into the XbaI site. Several plasmids were screened by sequencing to find variants in the correct orientation and insert number. The DNA encoding the protein to be exported (ADF silks or DH) is inserted between the HindIII and XbaI sites. The DH gene was received as a gift (Wendell Lim, UCSF). The three ADF silks were synthesized by DNA 2.0 and delivered in their pJ2 plasmid (pUC, Kan^R). Two pCASP derivatives were made eliminating the chaperone (-sicP) and chaperone/N-terminal SptP tag (-tag). The –sicP variant of pCASP was constructed by deletion PCR. The DH domain was fused to the N-

terminal SptP tag and driven by the *sicA* promoter and *sicA* ribosome binding site. The tagless (-tag) variant of pCASP was constructed by cloning the *sicA* promoter with a reverse primer including a start codon (ATG) and a HindIII site. This fragment was used to replace the *PsicA-sicP-sptP* fragment between the XhoI and HindIII restriction sites. A start codon and HindIII restriction site were added to the N-terminus of the DH domain (encoding the amino acids MKL). The DH domain is under the control of the *sicA* promoter and *sicA* ribosome binding site. –tag versions of pCASP were also constructed for all three silk monomers. These plasmids were constructed through a HindIII/XbaI digestion of the pCASP DH –tag plasmid and ligation of the silk gene into this location. This procedure also introduces the MKL amino acids at the N-terminus of the silks. Using deletion polymerase chain reaction we constructed a variant of pCASP removing the SptP secretion tag. Because the SicP and SptP genes overlap, the entire secretion tag could not be deleted and the first 5 amino acids (MLKYE) were left in the construct.

2.3.4. Construction of Alternative tag/chaperone Pairs.

Variations of pCASP were constructed containing different combinations of N-terminal tags and their cognate chaperones (details and references in Table 1). Each pair was cloned into the pCASP plasmid in place of the *sptP*/*sicA* pair under the control of the *sicA* promoter. The chaperone was amplified from genomic DNA and fused to the *sicA* promoter using SOEing PCR (Horton 1995). This piece was digested and ligated into pPROTet. Then, the tags were amplified, digested, and ligated into the plasmid. The resulting fusions were introduced into pPROTet between the XhoI and HindIII sites. An EcoRI site was introduced before the HindIII in the chaperone's reverse primer. Then

EcoRI/HindIII was used for all the tags. Finally the DH gene was introduced via digestion and ligation between the HindIII and XbaI sites.

2.3.5. Inducible Expression Plasmids.

The pTRC99a_Cm plasmid (ColE1 Origin, Cm^r) containing the IPTG-inducible P_{TRC} promoter was a gift from George Georgiou (UT-Austin). pTRC99a_Cm is a variation of the more common pTRC99a plasmid (Amersham-Pharmacia), only the Amp^r is replaced with Cm^r. A fragment containing *sicP-sptP*-DH-FLAG was amplified from the pCASP DH plasmid by PCR and ligated into the KpnI and XbaI sites of pTRC99a_Cm. For the growth and toxicity assays, inducible constructs were constructed for the three silk genes. The inducible silk variants were created by digestion (HindIII/XbaI) and gel purification of all silk genes followed by ligation into the backbone at the HindIII and XbaI sites to replace the DH gene.

2.3.6. Flow Cytometry and Construction of Fluorescent Reporters

Fluorescent reporters and the method for measuring promoter activity by flow cytometry was reported previously (Temme et al. 2008).

2.3.7. Coomassie Gels

A coomassie stained gel was run to compare supernatants to lysis controls. The DH band identity was confirmed by preparing an immuno-precipitated sample of DH protein. 10mL of DH containing supernatant was incubated with 6µL of mouse anti-FLAG

antibody for 3 hours at 4ºC. This was followed by 2 hour incubation with 0.5mL of washed (twice in 1mL phosphate buffered saline, pH 7.4 [PBS]) sepharose-bound Rat anti-Mouse IgG1 antibody (Zymed Cat#: 04-6141). The sepharose beads were collected by centrifugation at 1500xg for 5 minutes, transferred to a 1.5mL microfuge tube and washed 2x with PBS. Aliquots were prepared in sample buffer and run on a PAGE gel with coomassie staining. These gels (Lonza 10% Polyacrylamide Gels Cat#: 58102) were loaded with 10µl of protein precipitate (equivalent to 1.5mL 1× supernatant, procedure in 'Assay to detect lysis in secretion assay' section) and DH standard and run at 30mA for 1 hour. Staining was done by rinsing the gel in deionized water briefly followed by 1 hour in Coomassie blue stain with shaking (VWR model#100) followed by 8 hours in destain (10% v/v Ethanol, 10% v/v Glacial Acetic Acid, in deionized water) with Kimwipes (Kimberly Clark Cat#: 34155). For the positive lysis control, a 50mL culture of saturated wild-type *Salmonella* were by pelleted at 3500g 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 -80 ^{o}C. After thawing 0.2 mg/mL 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 (5000g for 5 minutes).

2.3.8. Secretion Assay and Extracellular Protein Collection

The SPI-1 T3SS is induced in high-salt LB (0.3M NaCl) and uninduced in LB-Lennox (L broth, 0.085M NaCl) (Tartera and Metcalf 1993). Cells were plated on L-broth agar plates from frozen stock and grown overnight. Single colonies were picked and grown 10 hours overnight in 5 ml liquid L broth. The overnights were diluted to an OD600 of 0.01

into 5mL fresh L-broth and grown for 2 hours at 250 rpm. The cultures were diluted a second time 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. The time point T = 0 is immediately after the second dilution step. Supernatants were harvested by spinning cultures at 3500g for 30 minutes followed by vacuum filtration through 0.45μm cellulose acetate filter unit (Corning Cat# 430314). The supernatant samples of proteins fused to the SptP N-terminal tag were used for the Western blots without concentration. The -tag silk monomers could not be visualized in 1x supernatant, so these samples were concentrated 30x by centrifugal filter (Amicron Cat# UFC801024) to quantitatively measure the background (Supplementary Information).

2.3.9. Preparation of Cell Pellet Samples

Cell pellets were collected to measure the amount of protein expressed, but not secreted. 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 0.2 mg/mL lysozyme (MP Biomedicals Cat#100834) and incubation for 15 minutes at room temperature followed by a 30 minute freeze/thaw cycle at -80^{o}C. Finally, the lysate was sonicated for 2 minutes in 1 second pulses with 30% amplitude. The resulting mixture was spun to remove insoluble debris at 3500g for 15 minutes. A 1mL sample of the soluble fraction was collected and frozen for quantitative western blot analysis.

2.3.10. Protease Cleavage

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.

2.3.11. Assay to Detect Lysis in Secretion Assay

Supernatant samples were precipitated in 10% trichloroacetic acid (TCA) (Acros Organics Cat#:152130010) at 4^{o}C for 1 hour. Precipitates were pelleted at 5000g for 30 minutes, washed with 1mL of PBS (pH 7.4) and precipitated again in 9mL of ice cold acetone. This mixture was re-pelleted at 5000g for 30 minutes and resuspended in 200μl of PBS. Samples were prepared in SDS sample buffer under reducing conditions and boiled for 3-5 minutes before PAGE analysis. MalE is the maltose binding protein located in the periplasm and is an indicator of periplamic leakage (Majander et al. 2005). The presence of MalE was determined via Western blot using an anti-MalE antibody (Genetex Cat# GTX20065).

2.3.12. Western Blots

The proteins to be detected were engineered to contain a C-terminal FLAG epitope tag for detection. Samples were prepared in SDS loading buffer under reducing conditions and run on 10% or 12% polyacrylimide gels (Lonza Biosciences). The gels were transferred to PVDF membranes (BioRad Cat#: 162-0177) and blocked for 1-8 hours in

TBST-5% non-fat milk (TBS-Tween20) with shaking. The anti-FLAG antibody (Sigma Cat# F3165) was used as the primary antibody. The anti-FLAG antibody was mixed 1:7,500 in TBST-milk and allowed to bind for 1 hour. After each antibody step, membranes were rinsed three times with deionized water followed by three 10 minute washes in TBST. The secondary antibody (Sheep anti-mouse HRP [Jackson ImmunoResearch Cat#: 515-035-003]) was added in TBST-milk at 1:5,000 and allowed to bind for one hour. Development was done using an enhanced chemiluminescent substrate for HRP (Pierce Cat#: 32209) and captured on film (Kodak: Cat#:178-8207).

2.3.13. Quantitative Western Blots

Quantitative western blots were run using an identical protocol as above until the secondary antibody step. Changes include replacing milk with Bovine Serum Albumin (BSA) for lower fluorescent background and using a Hybond low fluorescence PVDF membrane (Amersham Biosciences Cat#: RPN303LFP). We created a standard ladder of bacterial alkaline phosphatase with a C-terminal 1xFLAG tag (Sigma Cat# P-7457) of known concentration was run with each gel for construction of a standard curve. This standard curve was used in conjunction with our 3xFLAG measurements to convert between the differential sensitivities of 1x and 3x FLAG. A Cy-5 conjugated secondary antibody (GE Healthsciences Cat# PA45010) was used with a Typhoon 9400 variable mode imager to generate gel images (Gingrich et al. 2000; Biosciences 2001). Gel densitometry was done using ImageJ 1.41 (NIH). Protein yields were determined using the standard curve and linear point-to-point fitting interpolation. This data was then converted using the measured difference in sensitivity between 1x and 3x FLAG tags to

get the final yield number (Table 2, Supplementary Information). Raw data and conversion calculations for the quantitative westerns are provided in the Supplementary Information.

2.3.14. Construction, Purification and Testing of 3xFLAG BAP

No recombinant protein was available commercially with a 3x C-terminal FLAG tag and the extremely low secretion of the –tag or –chaperone samples required the extra sensitivity of the 3xFLAG (Hernan et al. 2000). In addition no quantitative measurements existed in the literature allowing us to convert the densitometry signal to an accurate measure of protein amount. To solve this problem, bacterial alkaline phosphatase (phoA, Genbank: 945041) was PCR amplified from genomic E.coli MG1655 DNA. A 6xHis tag was appended to the N-terminus and a 3xFLAG tag was appended to the C-terminus. The resulting construct was cloned into a pTRC expression vector and protein was purified on a cobalt affinity column according to the manufacturers instructions under denaturing conditions (Clontech Cat#: 635639). The eluted protein was analyzed by coomassie gel (Supplementary Information) and the eluted fraction was further purified. PhoA was subjected to secondary purification by size exclusion chromatography using an AKTA purifier UPC10W; Unicorn 5.11 software package (GE Healthcare Biosciences Corporation). Injections were loaded onto a Superdex 200 5/150 GL (GE Healthcare Biosciences Corporation Cat # 28-9065-61) previously equilibrated in 30 mM Tris HCl pH 8.0, 100 mM NaCl . The PhoA was eluted isocratically at 0.3 ml/minute in 0.5 ml fractions. The final product was checked for purity side by side with the 1xFLAG commercial sample on a silver stained gel (Biorad

Cat#: 161-0449) (Supplementary Information). The amount of protein in our 3x BAP standard was determined by the BCA assay (Pierce Cat#: 23235). Using this a defined number of femtomoles of 1x and 3x FLAG tagged protein were loaded on a 10% polyacrylimide gel (PAGE) and analyzed using the quantitative western blot method. The results were analyzed by gel densitometry and determined to be linear on a log-log plot (Supplementary Information). The power law best fit for each curve was determined from this information and used as a conversion factor between test samples run with 3xFLAG and standards on those gels with 1xFLAG.

2.4. 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. An example of this is the construction of cells that secrete enzyme cocktails to degrade plant biomass into simple sugars for the production of fuels and other products (Somerville 2007; Bayer et al. 2009). Finally, it could be used to expand the language of cell-cell communication through proteins and peptides, which is a frequently employed tool for cellular programming (Gerchman and Weiss 2004).

Here, we describe a control system that contains all of the necessary genetic parts for heterologous proteins to be secreted to the extracellular environment via the *Salmonella* SPI-1 T3SS (Figure 1). This system represents a novel route to secrete

proteins of biotechnological interest. 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 2) are as good or better than other methods of extracellular protein production (Choi and Lee 2004). Previously reported values range from 0.5 to 10 mg/L-hr 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-hr-OD for a protein the size of ADF2. 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. Because 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 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.

2.5. Supplemental Information for Chapter 2

2.5.1. Design and Synthesis of Silk Genes

The ADF amino acid sequences are highly enriched in a few amino acids. For example, over 80% of the ADF-3 sequence is comprised of only four amino acids (206 glycines,

109 glutamines, 104 alanines and 83 prolines out of 636 residues). The codons used to encode these amino acids in the naturally occurring gene are unevenly distributed and for glycine and proline the wild-type spider DNA sequence contains codons used poorly in *E. coli* (GGA and CCC respectively). We reasoned that by making the codon distribution more even, we could reduce the number of codons rarely used in highly expressed *E. coli* genes (Henaut and Danchin 1996) and increase the stability of the gene. To redesign of the spider genes for expression in *E. coli*, we therefore focused on redistribution of codons.

To design the synthetic gene encoding ADF-3, the coding sequence was first divided into three sections (A1 to G140, P141 to G314 and P315 to A636). Each of these sections was then back-translated independently using the GeneDesigner software package (Villalobos et al. 2006). The back translation algorithm initially selected a DNA sequence by choosing codons probabilistically using the frequencies found in *E. coli* class II genes (Henaut and Danchin 1996). The sequence was then evolved *in silico* to reduce the number of repeats: a codon would be selected at random and replaced by another codon for the same amino acid. If the change reduced the number or size of repeats it was accepted; otherwise, it was rejected. This process was repeated between 1,000 and 10,000 times.

The design criteria was applied to all three spider silk genes. Each synthetic gene encodes an amino acid sequence that exactly matches all of the available sequence information for the natural silk, corresponding to sequences between 294-635 amino acids. However, the synthetic genes share less than 29% codon identity with the natural DNA sequences. The improvement in codon usage is quantified in **Figure 2**A by

comparing the average codon frequency of the wild-type and synthetic silk genes. The reduction in repetitive DNA can be quantified as an increase in the information entropy of the nucleotide sequences encoding the repeat units (Figure 3B). A larger information entropy indicates that there is more underlying DNA diversity in the amino acid repeat units.

For synthesis, each gene was divided into six sections and a BsmBI site was added to each end of each section, so that after digestion with BsmBI each part would be flanked by 4 base 5' overhangs complementary to the overhangs of adjacent sections. The sections were synthesized by assembly of short oligonucleotides (Kodumal et al. 2004), then cleaved with BsmBI and ligated together with the vector.

The DNA sequences of the synthetic spider silk genes are shown, corresponding to the wild-type amino acid sequences. In each amino acid sequence, the highly conserved repetitive unit is colored. These repeat units are those used for the sequence entropies shown in Figure 3B of the main text.

> ADF-1 amino acid sequence (Genbank: AAC47008.1)

HESSYAAAMAASTRNSDFIRNMSYQMGRLLSNAGAITESTASSAASSASSTVTES

IRTYGPAAIFS

GAGAGAGVGVGGAGGYGQGYGAGA

GAGAGAGAGAGGAGGYGQGYGAGAAAAAGAGAGAAG

GYGGGSGAGAGGAGGYGQGYGAGSGAGAGAAAAAGASAGAAGGYGG

GAGVGAGAGAGAAGGYGQSYGSGAGAGAGA

GAAAAAGAGARAAGGYGGGYGAGA

GAGAGAAASAGASGGYGGGYGGGAGAGAVAGASAGSYGGAVNRLSSAGAAS

RVSSNVAAIASAGAAALPNVISNIYSGVLSSGVSSSEALIQALLEVISALIHVLGSA

SIGNVSSVGVNSALNAVQNAVGAYAG

> Synthetic ADF-1 gene

CATGAATCTTCCTATGCTGCTGCAATGGCTGCTTCTACTCGTAATTCTGATTTT

ATCCGTAACATGAGCTACCAGATGGGTCGTCTGCTGAGCAACGCCGGTGCCA

TTACCGAATCTACTGCAAGCAGCGCGGCTTCCAGCGCGTCCTCCACCGTTACC

GAGTCTATTCGCACGTATGGCCCGGCTGCGATCTTTTCTGGTGCGGGCGCTGG

CGCAGGCGTGGGTGTAGGTGGTGCCGGTGGTTACGGCCAGGGCTACGGTGCA

GGCGCAGGTGCTGGTGCGGGCGCCGGTGCGGGTGCTGGTGGCGCGGGTGGCT

ACGGTCAGGGCTACGGTGCGGGTGCTGCCGCGGCGGCTGGCGCAGGCGCAG

GTGCAGCGGGCGGTTACGGCGGTGGTTCTGGTGCTGGCGCTGGTGGTGCTGG

TGGTTATGGCCAGGGTTACGGTGCAGGTTCTGGCGCGGGTGCGGGCGCTGCT

GCGGCAGCTGGCGCATCCGCTGGTGCTGCTGGCGGCTATGGCGGTGGCGCAG

GTGTTGGTGCAGGTGCGGGCGCGGGTGCGGCTGGTGGCTATGGCCAGAGCTA

TGGCAGCGGTGCTGGCGCAGGTGCGGGTGCTGGTGCGGCGGCTGCAGCTGGC

GCTGGCGCACGTGCAGCGGGTGGCTACGGTGGTGGTTACGGCGCAGGCGCGG

GCGCCGGTGCTGGCGCCGCTGCTTCCGCTGGTGCCTCCGGTGGCTACGGTGG

CGGTTACGGCGGTGGCGCGGGTGCAGGCGCCGTAGCTGGTGCGTCCGCGGGT

TCTTACGGCGGTGCGGTTAACCGTCTGTCTAGCGCAGGCGCGGCATCTCGTGT

TTCCAGCAACGTGGCTGCCATCGCGTCTGCGGGTGCGGCTGCGCTGCCGAAC

GTAATCTCTAACATTTATTCTGGTGTGCTGTCCTCTGGTGTGTCTTCTTCTGAG

GCGCTGATCCAGGCTCTGCTGGAAGTCATCTCTGCACTGATCCACGTGCTGGG

TTCTGCCTCCATCGGTAACGTGTCTTCCGTTGGCGTTAACAGCGCACTGAATG

CAGTGCAGAACGCCGTCGGCGCGTACGCTGGT

> ADF-2 amino acid sequence (Genbank: AAC47009.1)

GSQGAGGAGQGGYGAG

GGGAAAAAAAAVGAGGGGQGGLGSGGAGQGYGAGLGGQ

GGASAAAAAAGGQGGGQGGGQGGYGGLGSQGAGGAGQLGYGAG

QESAAAAAAAAGGAGGGGQGGLGAGGAGQGYGAAGLGGQGGAGQ

GGGSGAAAAAGGQGGGQGGYGGLG

PQGAGGAGQGGYGGGSLQYGGQGQAQAAAASAAASRLSSPSAAARVSSAVSL

VSNGGPTSPAALSSSISNVVSQISASNPGLSGCDILVQALLEIISALVHILGSANIGP

VNSSSAGQSASIVGQSVYRALS

> Synthetic ADF-2 gene

GGTAGCCAAGGCGCAGGTGGTGCAGGTCAAGGTGGTTATGGTGCAGGCGGC

GGTGGCGCTGCGGCAGCTGCTGCTGCAGCGGTAGGCGCGGGTGGCGGTGGTC

AGGGCGGCCTGGGTTCCGGCGGTGCGGGCCAGGGTTACGGCGCAGGCCTGG

GCGGTCAAGGTGGCGCATCTGCGGCGGCTGCTGCGGCTGGTGGCCAGGGCGG

TCAGGGTGGCCAAGGTGGCTATGGCGGTCTGGGTTCTCAGGGCGCAGGCGGT

GCTGGTCAGCTGGGCTATGGTGCAGGTCAGGAATCTGCAGCGGCTGCCGCTG

CCGCAGCGGGCGGCGCTGGTGGCGGTGGTCAGGGCGGCCTGGGTGCGGGTG

GCGCTGGCCAAGGTTACGGTGCCGCTGGCCTGGGCGGTCAGGGTGGTGCGGG

CCAGGGCGGCGGCTCTGGCGCGGCGGCTGCGGCCGGTGGTCAAGGTGGTCAG

GGCGGCTATGGTGGCCTGGGCCCGCAAGGCGCGGGTGGTGCGGGCCAGGGT

GGCTACGGTGGTGGTTCCCTGCAATACGGCGGTCAGGGTCAGGCTCAGGCAG

CTGCGGCATCCGCGGCGGCATCCCGCCTGTCTTCCCCATCCGCAGCGGCACGT

GTGTCTTCCGCTGTATCTCTGGTATCCAACGGCGGTCCGACCAGCCCGGCGGC

ACTGAGCTCTAGCATTTCCAACGTGGTATCTCAGATCTCTGCAAGCAACCCAG

GCCTGTCTGGTTGCGATATCCTGGTTCAAGCCCTGCTGGAAATTATCTCTGCG

CTGGTTCACATCCTGGGTTCTGCCAACATCGGCCCGGTTAACTCTAGCTCCGC

CGGTCAGTCCGCATCCATTGTAGGTCAATCCGTATACCGCGCTCTGTCT

> ADF-3 amino acid sequence (Genbank: AAC47010.1)

ARAGSGQQGPGQQGPGQQGPGQQGPYGPG

ASAAAAAAGGYGPGSGQQGPSQQGPGQQGPGGQGPYGPG

ASAAAAAAGGYGPGSGQQGPGGQGPYGPG

SSAAAAAAGGNGPGSGQQGAGQQGPGQQGPG

ASAAAAAAGGYGPGSGQQGPGQQGPGGQGPYGPG

ASAAAAAAGGYGPGSGQGPGQQGPGGQGPYGPG

ASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGGQGPYGPG

ASAASAASGGYGPGSGQQGPGQQGPGQQGPYGPG

ASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGGQGPYGPG

ASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQGPGQQGPGQ

QGPGQQGPGGQGAYGPG

ASAAAGAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQGPGQQGPGQ

QGPYGPG

ASAAAAAAGGYGPGSGQQGPGQQGPGQQG

PGGQGPYGPGAASAAVSVGGYGPQSSSVPVASAVASRLSSPAASSRVSSAVSSLV

SSGPTKHAALSNTISSVVSQVSASNPGLSGCDVLVQALLEVVSALVSILGSSSIGQI

NYGASAQYTQMVGQSVAQALA

> Synthetic ADF-3 gene

GCCCGCGCGGGGTCAGGCCAGCAGGGACCAGGTCAACAGGGCCCGGGCCAAC

AAGGCCCGGGTCAACAGGGTCCGTACGGTCCGGGTGCCAGCGCGGCGGCCGCG

GCCGCAGGAGGGTATGGCCCTGGTAGCGGCCAACAGGGTCCGAGCCAGCAAG

GCCCGGGCCAGCAAGGGCCGGGGGGGCCAGGGGCCCTACGGCCCTGGTGCGTCA

GCTGCCGCAGCCGCAGCTGGCGGTTATGGCCCGGGGTCAGGTCAGCAAGGGCC

AGGCGGTCAAGGTCCTTACGGGCCAGGCAGTAGTGCGGCAGCGGCTGCTGCCG

GTGGTAACGGCCCGGGGTCGGGCCAGCAAGGGGCGGGACAGCAGGGTCCAGG

CCAACAAGGCCCCGGTGCGTCCGCAGCGGCGGCGGCCGCTGGTGGCTATGGCC

CGGGTTCAGGCCAGCAGGGCCCGGGGCAGCAGGGCCCGGGTGGACAGGGTCC

GTATGGCCCGGGGGCCAGTGCAGCGGCCGCGGCTGCTGGGGGCTATGGCCCTG

GCTCAGGTCAGGGTCCGGGTCAACAAGGACCCGGCGGTCAAGGACCGTATGGC

CCGGGTGCGTCCGCGGCGGCTGCGGCGGCTGGAGGCTATGGTCCGGGAAGTGG

CCAACAGGGCCCTGGACAGCAGGGTCCGGGTCAGCAGGGACCCGGTGGACAG

GGCCCGTATGGGCCAGGCGCCTCTGCCGCAGCGGCGGCCGCAGGTGGGTATGG

ACCGGGGTACGGCCAGCAGGGTCCTGGTCAGCAGGGACCGGGCGGCCAGGGC

CCTTACGGCCCCGGCGCGTCAGCTGCAAGCGCTGCCTCGGGTGGCTACGGCCC

GGGTTCCGGTCAGCAGGGCCCGGGACAGCAGGGTCCGGGTCAGCAGGGACCGT

ATGGTCCGGGAGCTTCTGCTGCTGCCGCCGCGGCGGGTGGTTATGGACCCGGC

AGTGGCCAACAAGGTCCGGGGCAGCAGGGTCCAGGTCAGCAGGGCCCAGGAC

AGCAGGGCCCTGGTGGCCAAGGACCGTACGGTCCCGGCGCAAGTGCGGCCGCT

GCAGCTGCCGGAGGCTACGGTCCAGGTAGTGGACAGCAAGGACCGGGTCAGC

AGGGCCCCGGTCAACAGGGGCCGGGCCAGCAAGGCCCCGGGCAGCAGGGACC

TGGGCAGCAGGGTCCCGGGCAGCAAGGTCCTGGGCAACAGGGTCCGGGACAG

CAAGGCCCTGGCGGCCAGGGTGCGTATGGGCCTGGTGCATCTGCCGCGGCGGG

CGCCGCGGGTGGGTACGGGCCGGGGAGCGGCCAGCAAGGTCCGGGCCAACAG

GGCCCCGGACAACAGGGTCCTGGCCAGCAAGGACCTGGCCAGCAGGGGCCGG

GACAACAAGGGCCCGGCCAACAAGGCCCAGGGCAACAAGGCCCGTACGGCCC

TGGGGCCTCGGCAGCCGCGGCAGCGGCCGGCGGCTATGGCCCGGGCAGTGGTC

AACAAGGTCCAGGCCAACAGGGCCCAGGGCAGCAGGGTCCGGGGGGTCAAGG

TCCGTACGGACCGGGTGCCGCCTCGGCAGCGGTGAGTGTAGGCGGCTACGGAC

CTCAAAGCTCCTCTGTGCCAGTCGCCAGTGCGGTGGCTAGCCGTCTGTCTAGCC

CCGCCGCCAGCAGTCGTGTCAGCTCAGCCGTGTCGTCTTTAGTATCATCAGGAC

CGACTAAACACGCAGCCTTGTCAAACACCATTAGCAGCGTTGTCTCTCAGGTGT

CAGCGAGTAACCCGGGGCTGTCGGGTTGCGACGTCCTGGTACAGGCCCTGCTG

GAAGTGGTGAGCGCCCTCGTGTCTATTCTGGGTTCTAGTTCCATTGGCCAGATT

AACTATGGGGCGAGTGCGCAATACACCCAAATGGTCGGACAATCTGTTGCGCA

GGCACTGGCG

2.5.2. Comparison With pTRC Promoter

A timecourse was run to compare the *psicA* promoter to the induction of the commonly used *pTRC*. A fluorescent reporter was created by fusing green fluorescent protein to the pTRC99a vector with the Ampicillin marker switched for a Chloramphenicol marker for consistency with the pCASP vector. Samples were collected every half hour for five hours and analyzed by flow cytometry in a BD LSRII (Figure 5).

All strains were grown using the secretion assay protocol (Materials and Methods) and *pTRC* samples had the noted amount of IPTG added at T=0 (when cells are shifted into SPI-1 inducing media). From this data, we concluded that *psicA* has a magnitude of induction approximately equal to 10μM IPTG induction of *pTRC*.

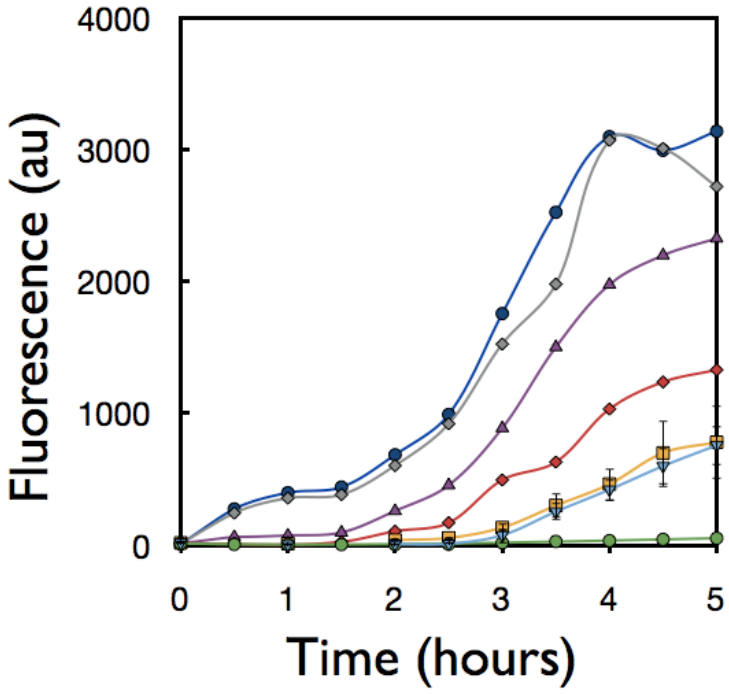


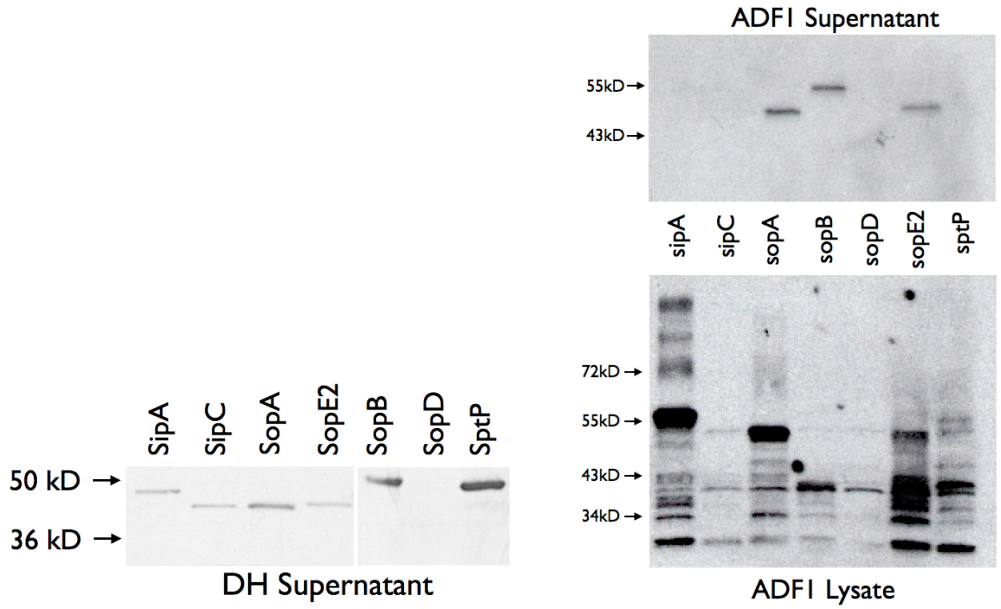
Figure 5

Curves showing fluorescence of GFP driven by a *pTRC* promoter over time for various IPTG inducer concentrations in *Salmonella* SL1344. IPTG was varied from 0μM (green circles), 10μM (yellow squares), 30μM (red diamonds), 100μM (purple triangles), 500μM (grey diamonds), and 1mM (blue circles). A curve showing *psicA* GFP (blue triangles) is shown for reference. Error bars for *psicA* and 10μM represent ±1 SD of three colonies measured on different days.

2.5.3. Tag / Chaperone Comparison

To select the best tag / chaperone pair for each protein a series of vectors were created with known tags and chaperones inserted between the XhoI and HindIII restriction sites on pCASP, while preserving the *psicA* promoter and TEV cleavage site. Each of these vectors was then digested with HindIII and XbaI restriction endonucleases and into this

site we cloned the DH, ADF1, ADF2, and ADF3 genes. All constructs were transformed into *Salmonella* SL1344 and run in a secretion assay (Figure 6). The highest secreted amount was chosen as the best construct. In the case of the secreted amount being equal the highest ratio of secreted protein to expressed protein was used to determine the best chaperone / tag pair.



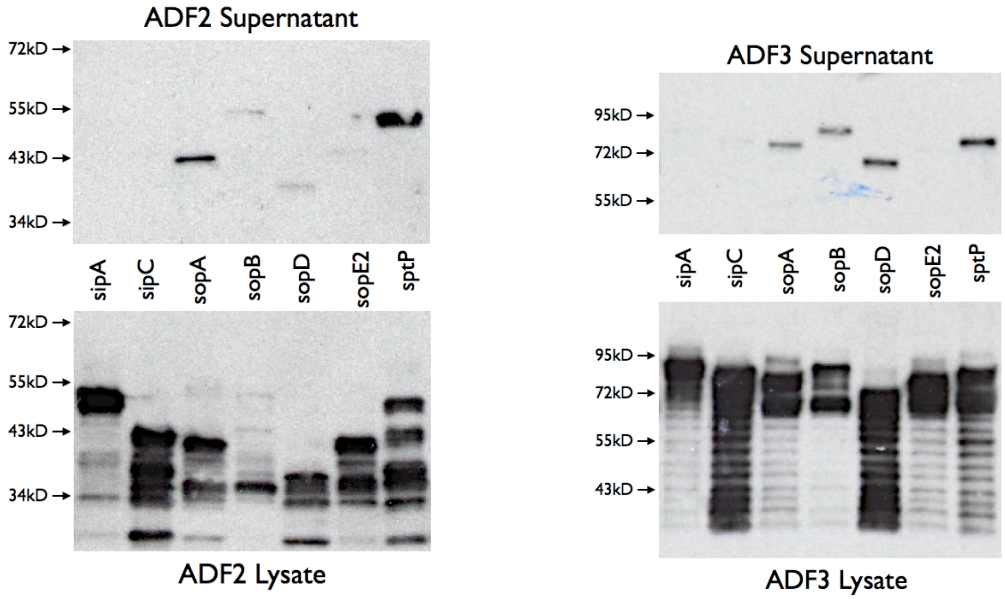


Figure 6

Testing the effect of tag/chaperone combinations versus the amount of secreted protein or efficiency. The tag used in each lane is labeled (the cognate chaperone is listed in Table 1). For DH only the amount of secreted protein was measured as a test. For (B-D) all silks were compared with known tag/chaperone combinations. The supernatant is on the top gel image and the corresponding lysate samples are on the bottom. Using this data, the highest amount of protein secreted was selected as the best tag/chaperone pair.

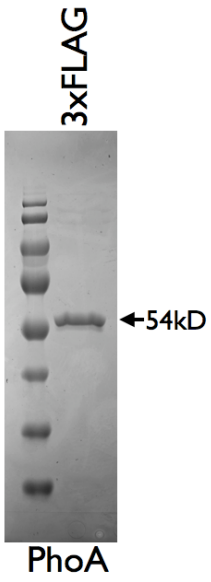
2.5.4. 3x FLAG PhoA Purification and Standard Calibration

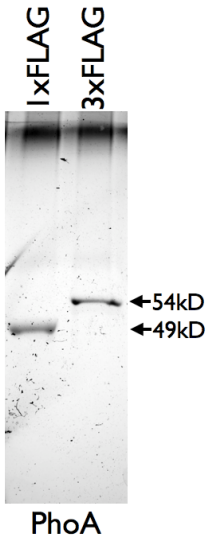
In determining a quantitative measure of protein secretion we chose the method of quantitative western blotting. All secreted proteins were genetically engineered to contain a 3x FLAG $(DYKDDDDK)_3$ affinity tag on the C-terminus. This tag was detected with the anti-FLAG antibody (Sigma Cat# F3165). The 3x FLAG is required for the greatly improved sensitivity and allows detection of the very dilute (-) tag supernatants that are difficult to detect even at 30x concentration.

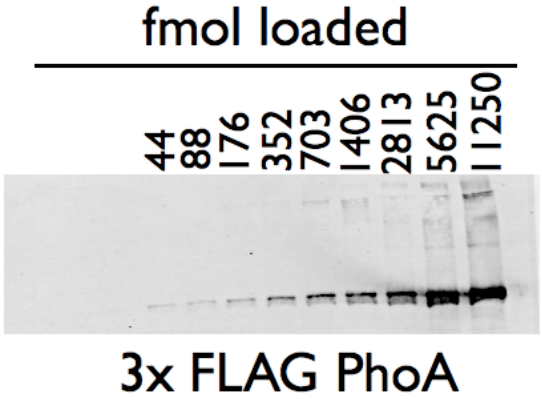
In order to carry out quantitative western blotting a standard ladder of known protein amount is run concurrently with samples of interest. The standard ladder is used to create a standard curve and interpolate the amount of protein in the sample. The only available commercial sample of FLAG tagged protein is the 1xFLAG C-terminal fusion to bacterial alkaline phosphatase (Sigma Cat# P-7457). We carried out experiments to determine and correct for the difference in sensitivity between 1x and 3x FLAG tagged proteins.

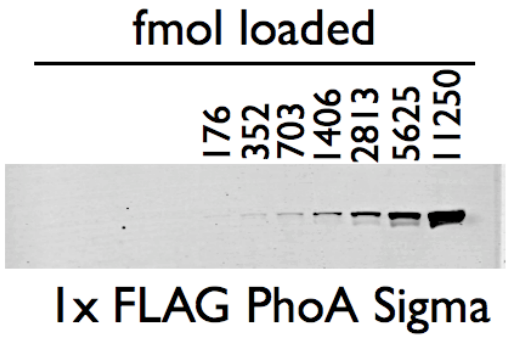
Bacterial Alkaline Phosphatase (PhoA, Genbank: 945041) was cloned and purified with a 3xFLAG tag attached (Materials and Methods) (Figure 7A). The final eluted protein was quantified using the BCA protein quantitation assay (Pierce Cat#: 23235) and diluted to the same concentration as the commercial 1xFLAG (10,000 femtomoles per microliter). This product was run side-by-side with the commercial 1xFLAG PhoA on a 10% PAGE gel and stained with silver stain (Biorad Cat#: 161-0449) (Figure 7B).

The 1x and 3x FLAG samples were run in identical amounts on separate 10% PAGE gels side by side (Figure 7C and Figure 7D) and analyzed by quantitative western when developed in the same solutions using the quantitative western blot method (Materials and Methods). The gels were analyzed for signal intensity relative to femtomoles of protein loaded and the two standard curves were solved to give the relative sensitivity of the FLAG tags (Figure 7E). The best fit on the log-log plot was a power law and the equation of this fit was used to correct quantitative western blot data for 3xFLAG samples measured against 1xFLAG standard ladders.


3xFLAG
54kD
PhoA


1xFLAG
3xFLAG
54kD
49kD
PhoA


fmol loaded
44
88
176
352
703
1406
2813
5625
11250
3x FLAG PhoA


fmol loaded
176
352
703
1406
2813
5625
11250
1x FLAG PhoA Sigma

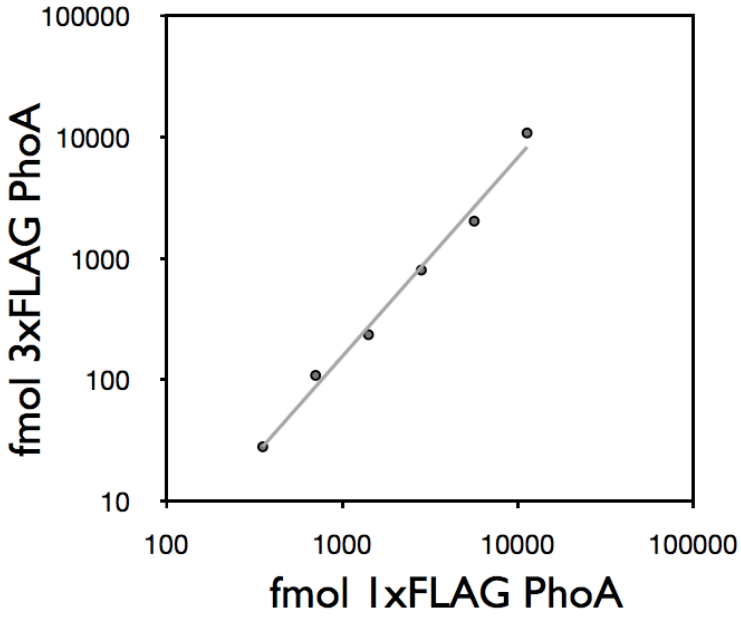


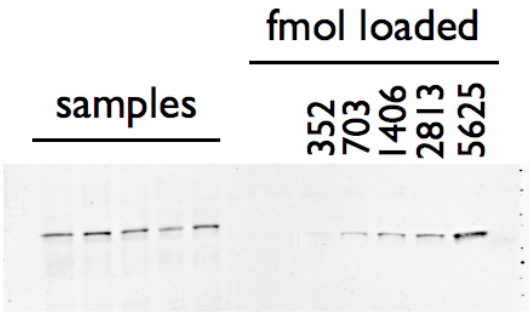
Figure 7

Creation of a 3xFLAG standard and determining the difference in sensitivity. **(A).** A coomassie brilliant blue stained PAGE gel of 6xHis PhoA 3xFLAG after purification with His-tag affinity purification. **(B).** A silver stained PAGE gel of the commercial PhoA 1xFLAG (1xFLAG) and PhoA 3xFLAG (3xFLAG) after size exclusion purification and quantitation by BCA assay. Equal amounts of 1xFLAG and 3xFLAG were loaded on this gel and can be clearly seen. **(C) & (D).** A serial dilution of 1xFLAG PhoA and 3xFLAG PhoA respectively are run using the quantitative western blotting method (Materials and Methods). **(E).** A log-log plot showing the relative sensitivities of 1x and 3x FLAG PhoA and a power-law best-fit curve [Eqn: $y=0.0019x^{1.6391}$]. This curve was used as a calibration to compensate 3xFLAG samples run against 1xFLAG standard curves to adjust for the sensitivity difference. The gel images shown were adjusted for contrast to enhance reproduction for publication. During analysis the raw data was unaltered.

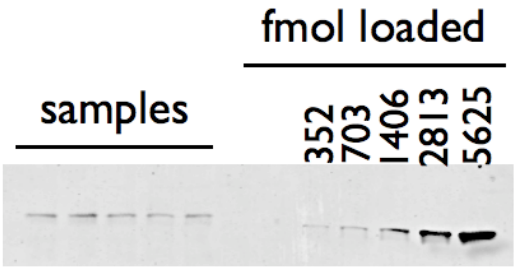
2.5.5. Quantifying Silk Yield

All silk secretion samples were analyzed by quantitative western blot (Materials and Methods) to determine the amount of protein expressed and secreted with and without the secretion tag (Table 2). A set of five individual colonies were picked in parallel, samples

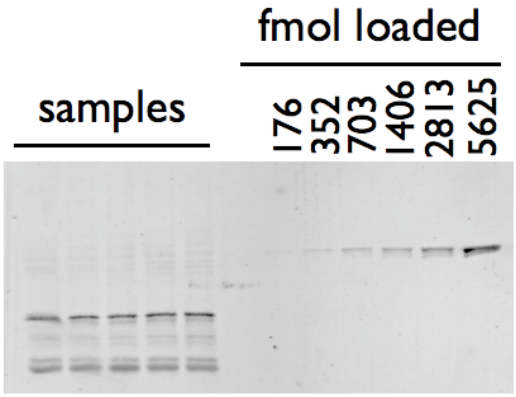
were grown for each tagged and (-) tag silk and prepared as described. All of the quantitative western blots used to generate the yield numbers are shown (Figure 8A-L). Gels were analyzed in ImageJ for oversaturation (defined as pixels reaching maximum intensity) and points meeting this criteria were not considered.



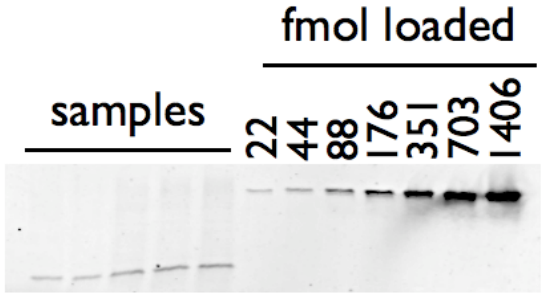
SopB ADF1 Lyse, Concentration: 1x



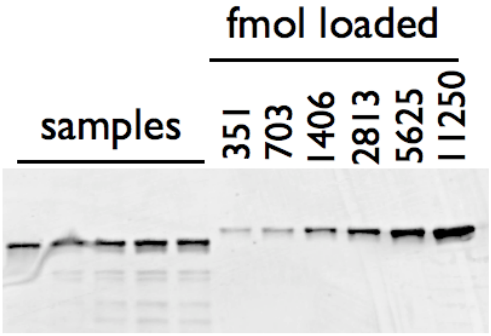
SopB ADF1 Supe, Concentration: 1x



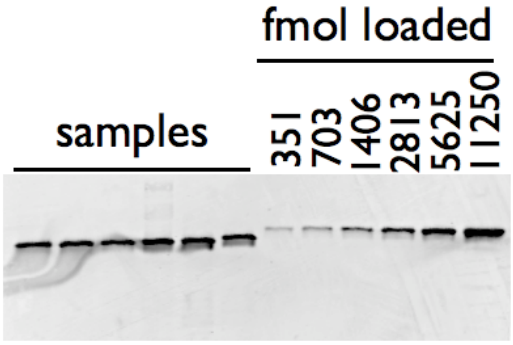
-Tag ADF1 Lyse, Concentration: 1x



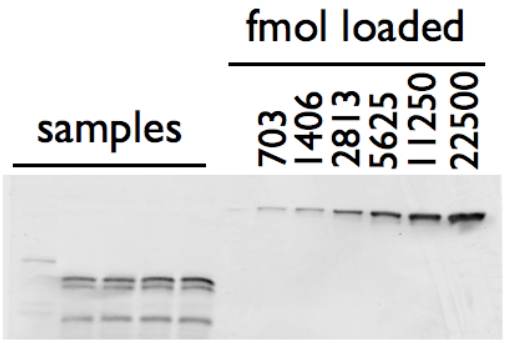
-Tag ADF1 Supe, Concentration: 30x



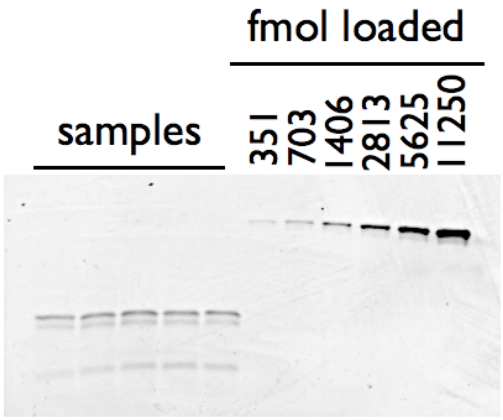
SptP ADF2 Lyse, Concentration: 0.05x



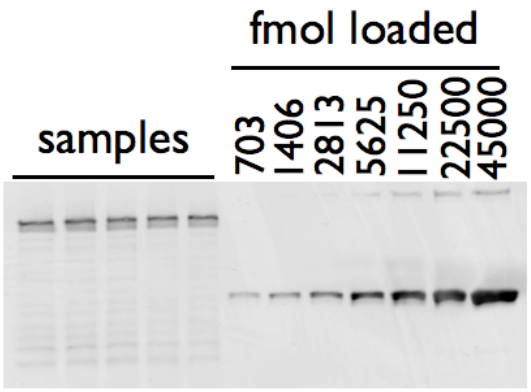
SptP ADF2 Supe, Concentration: 1x



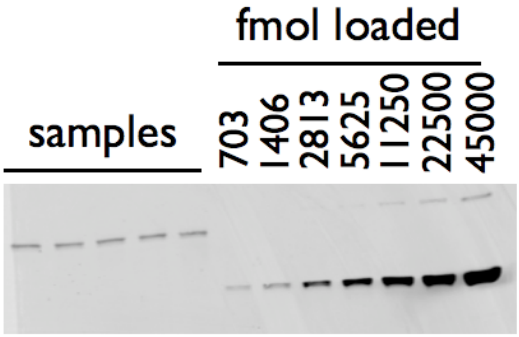
-Tag ADF2 Lyse, Concentration: 0.05x



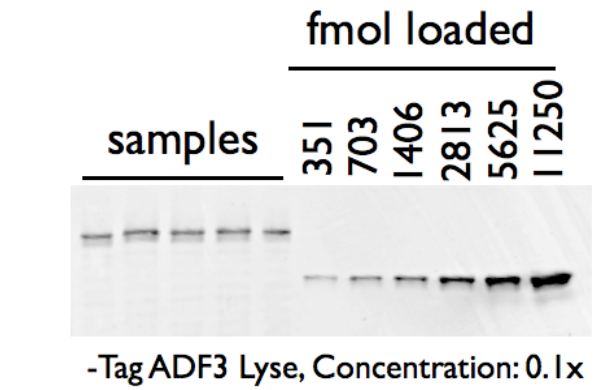
-Tag ADF2 Supe, Concentration: 30x



SptP ADF3 Lyse, Concentration: 0.1x



SptP ADF3 Supe, Concentration: 1x



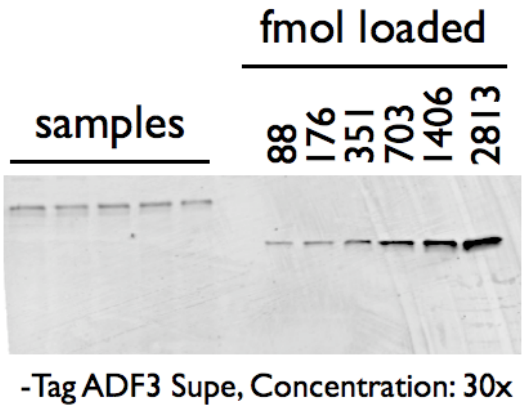


Figure 8

Quantitative western blot scans. These are the images analyzed to determine the amount of protein in the supernatant and lysate for tagged and (-) tag silk. Each gel has a standard curve of 1xFLAG BAP with a defined number of femtomoles loaded and several sample lanes. The concentration of the samples in each lane is noted below the gel image. Because the dynamic range of the detector is fairly limited, and samples concentration varied widely, the samples were diluted or concentrated to fall within the range of the ladder. The gel images shown were adjusted for contrast to enhance reproduction for publication. During analysis the raw data was unaltered when analyzed.

2.5.6. Additional Experiments with T3SS Knockouts

Two knockout strains were obtained from the Galan lab and secretion experiments were performed on those strains according to the published protocol (Eichelberg and Galan 1999; Lee and Galan 2004). Secretion assays on the Δ*invA* (SB136) and Δ*flaGHI* (SB181) and wild-type (*Salmonella typhimurium* SL1344) *Salmonella* strains carrying the IPTG-inducible DH plasmid were carried out as follows (Figure 9). Colonies were streaked from frozen stock, picked into 5mL of SPI-1 inducing media, and grown for 10-12 hours at 37 ^{o}C and 250rpm. All cultures were diluted ~1:50 into 50mL of fresh inducing media (0.3 M NaCl LB) in a 250mL non-baffled flask supplemented with antibiotic and 100μM IPTG. The dilutions were corrected for OD_{600} so that all 50mL cultures had the same starting OD_{600}. The cultures were grown for 4 hours at 37 ^{o}C and 160rpm. Supernatants were harvested by spinning cultures at 3500g for 30 minutes followed by vacuum filtration through a 0.45μm cellulose acetate filter unit (Corning Cat# 430314). At this point sample proteins were precipitated in 10% trichloroacetic acid (TCA) for 1 hour and pelleted at 5000g for 30 minutes. Pellets were washed with 9mL of cold acetone and re-pelleted at 5000g for 30 minutes. The acetone supernatant was decanted and the resulting pellets were air dried for 2 hours. The final pellet was resuspended in 200μL of PBS and frozen for later PAGE analysis. Western blots were then performed on the samples as described (Materials and Methods).

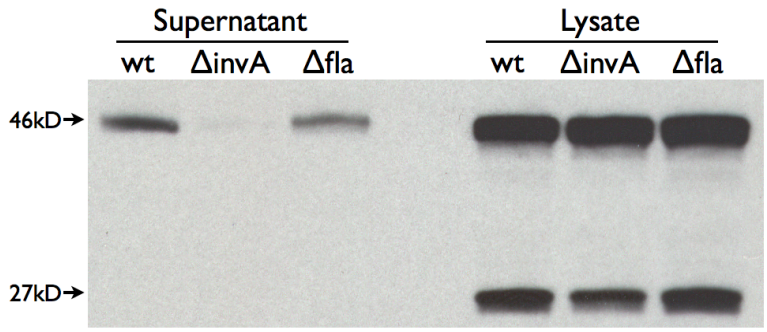


Figure 9

Samples are shown from the supernatant and lysate and shown for the wild-type (wt) and SPI-1 (ΔinvA) and flagella (ΔflaGHI) knockout strains.

2.5.7. Circuit Component Effects on Growth

A series of growth curves were run to control for toxicity of the various circuit components and secretion payloads. This was done by growing up 3-5 separate colonies in a secretion assay (Materials and Methods) and measuring OD_{600} at one-hour intervals (Figure 10). No obvious difference was detected for any version of the circuit (pCASP ADF1, pCASP ADF2, pCASP ADF3, pProTet, *psicA* SicP SptP DH, *psicA* SicP DH).

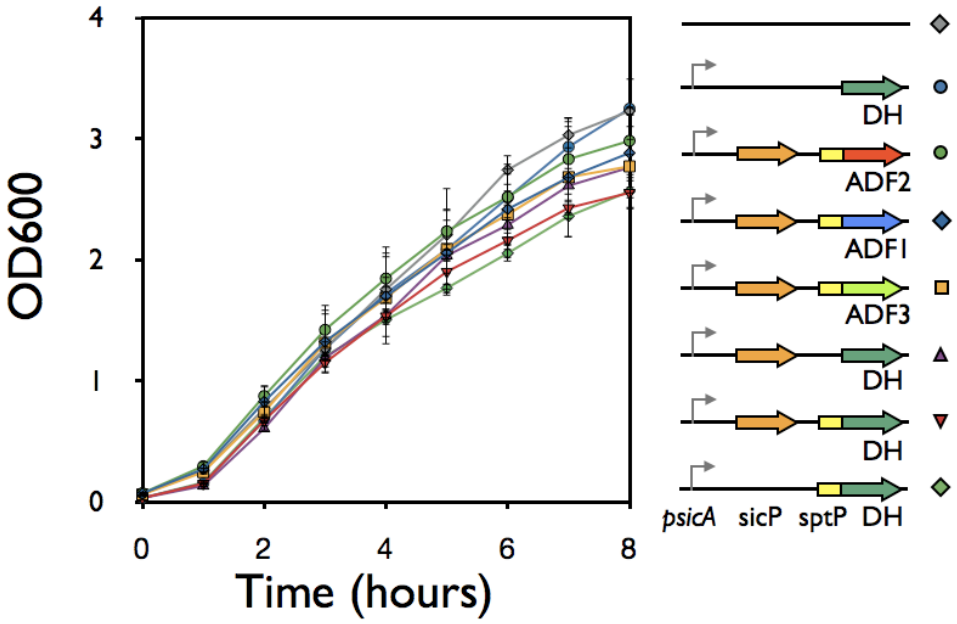


Figure 10

A series of growth controls testing circuit components for growth defects. Five fresh colonies of the circuit variants were picked and grown in a secretion assay (Materials and Methods) and OD_{600} was measured every hour. T=0 represents when the cells were shifted into inducing media. Error bars represent ±1 SD of the data. Samples shown are pCASP ADF1 (blue diamonds), pCASP ADF2 (green circles), pCASP ADF3 (yellow squares), pCASP DH (red triangles), *psicA* SicP DH (purple triangles), *psicA* SptP DH (green diamonds), *psicA* DH (blue circles) and a control plasmid (grey diamonds).

Chapter 3 Controlling Type III Secretion with Synthetic Circuits

3.1. Introduction

Controlling complex cellular functions, such as protein secretion, using synthetic genetic circuits is a critical goal for the field of synthetic biology. The type III secretion system (T3SS) is a unique molecular machine amongst eubacteria and is found in numerous gram-negative pathogens (Galan and Collmer 1999). The core feature is the ability to export proteins in a semi-unfolded state through both membranes of bacteria and into the culture media or host cells in a single step (Feldman et al. 2002; Lee and Schneewind 2002; Akeda and Galan 2005). The most recognizable feature of the T3SS is the needle complex, a molecular machine resembling a hypodermic needle (Marlovits et al. 2004; Yip et al. 2005). The ability to control protein secretion is of particular interest for biotechnology applications where the desired protein is harmful to the cell or a substrate cannot diffuse through the lipid membrane (Majander et al. 2005; Sorensen and Mortensen 2005; Lee et al. 2006; Zhang et al. 2006). Proteins secretion into the culture media also simplifies purification procedures and acts as a first step in separating the product of interest from contaminating proteins, nucleic acids, and lipids. The

engineering of whole organisms requires access to functional genetic modules that allow the export of proteins into the extracellular environment for function (ie: cellulases) or into host cells to reprogram cellular functionality (Zhang et al. 2006; Purnick and Weiss 2009).

In *Salmonella* the T3SS is contained in a single 34 kilobase gene cluster, *Salmonella* Pathogeneity Island 1 (SPI-1), and is a well-studied model system (Altier 2005). The system is highly induced under laboratory conditions by increasing culture broth osmolarity and decreasing dissolved oxygen (Tartera and Metcalf 1993). The genetics of this system are well described in the literature and recent work has shown the promise of SPI-1 as both a biotechnology tool and a therapeutic agent (Russmann et al. 1998; Widmaier et al. 2009). Proteins secreted by the T3SS contain an N-terminal peptide tag, which binds a specific secretion chaperone protein (Stebbins and Galan 2001). This complex directs proteins to the SPI-1 T3SS for secretion at the proper time and location (Lee and Galan 2004).

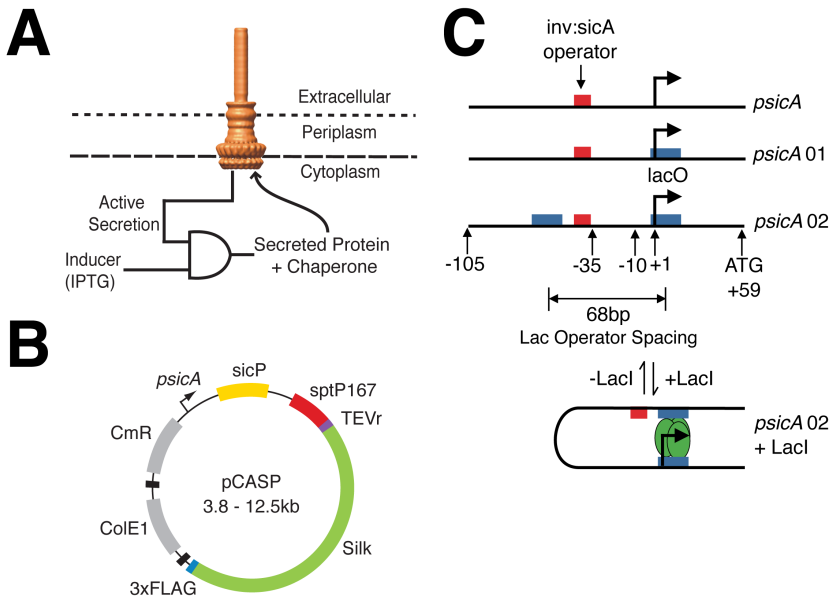


Figure 11

An overview of the system design described here is presented. **A.** The needle structure of *Salmonella* Type 3 Secretion exports proteins through both membranes and into the extracellular environment. Secreted spider silk proteins in this study are controlled by a transcriptional logical AND gate with inputs of natural system feedback (active secretion) and synthetic inducer added to the culture media (IPTG). When both signals are present secreted protein and chaperone are produced and exported. Needle image is a cryo-EM scan from (Marlovits et al. 2004) **B.** A map of the plasmid, pCASP, used in this study to generate the AND gate and control expression of secreted spider silk protein. Various promoters are generated to replace *psicA* with additional functionality. The DNA size range shown encompasses the range of plasmid sizes in this study with different size silk genes. The black features represent transcriptional terminators. **C.** Three promoters used in this study are shown with critical features marked (sequences available in supplemental information). The *psicA*02 promoter has two lac operator binding sites with appropriate spacing to allow DNA looping. This function allows for tighter repression of the promoter in the presence of LacI protein.

The ability to integrate native regulation and non-native control represents a major step towards engineering large genetic systems. Current genetic circuits for SPI-1

regulated expression integrate natural feedback from the secretion system to prevent protein expression until the secretion system is functional and ready to export protein (Widmaier et al. 2009). This feedback is integrated into the *psicA* promoter, the natural control point for an operon of secreted proteins (Darwin and Miller 1999; Darwin and Miller 2000; Tucker and Galan 2000; Darwin and Miller 2001). The ability to exert additional control upon *psicA* while conserving the SPI-1 feedback mechanisms is presented here (Figure 11A). This control enables experiments to answer outstanding questions the dynamics of protein expression and secretion as well as the biophysical limitations of secreted proteins (Temme et al. 2008).

Amongst the outstanding questions about SPI-1 as a biotechnology tool is the profile of secretion throughput over time. The natural context of SPI-1 is to transiently secrete effector proteins into mammalian host cells to aid in bacterium survival, at which point SPI-1 is turned off and an analogous T3SS, *Salmonella* Pathogeneity Island 2, is turned on (Kaniga et al. 1995; Lundberg et al. 1999; Altier 2005; Schlumberger et al. 2005). For protein expression applications presented here SPI-1 is used to secrete proteins into the culture broth. In this scenario it is unclear how secretion behaves over time. The ability to control discrete aspects of natural SPI-1 regulation with synthetic genetic circuits could answer this question. Here we add tight external control of the *psicA* promoter in SPI-1 to understand secretion into the culture media over time.

Methods to control gene expression using transcriptional control are well documented as essential components of the molecular biology revolution (Guzman et al. 1995; Lutz and Bujard 1997). The use of regulatory components from the *E.coli* lactose utilization operon (lac) has given rise to a ubiquitous toolkit of commercially available

plasmids and regulatory systems for controlling heterologous protein expression (Jacob and Monod 1961). One mechanism identified in the literature as LacO looping has yet to be examined for use in synthetic regulation (Besse et al. 1986; Mossing and Record 1986; Law et al. 1993; Muller et al. 1996; Mehta and Kahn 1999). In the lac system the LacI repressor tetramer creates symmetrical binding sites for the lac operator DNA sequence and can simultaneously bind two operators at once. This effect is dependent upon the concentration of LacI in the cell, the affinity for the operator-binding site, and most importantly the appropriate spacing between lac operator sites (Muller et al. 1996). When looping occurs the entire complex sterically occludes the binding of RNA polymerase to the -10 and -35 hexamers and preventing transcription initiation.

This system is well described as a natural regulatory mechanism but has remained untapped for synthetic regulation. In this paper we use LacO looping to create transcriptional logic that connects to the natural SPI-1 regulation (Figure 11A, Figure 11C). This system creates a series of hybrid *psicA* promoters that have different activation dynamics. The tight control of secreted protein expression allows an examination of SPI-1 function over time presented here.

Using SPI-1 as a tool for protein expression requires a refined understanding of which biophysical properties of substrates are critical limitations to protein export. Current work has focused on the properties of the endogenous effector proteins secreted by SPI-1 (Collazo and Galan 1996). The heterologous protein substrate chosen here is silk protein. Silk is a high value protein based material produced by many arthropods with each specific silk having tailored mechanical properties (ie: tensile strength, elasticity) to a specific physiological role such as prey capture or egg sack protection

(Lazaris et al. 2002; Kluge et al. 2008; Sutherland et al. 2009). The remarkable properties of silk protein are a result of primary amino acid sequences that are notoriously difficult manipulate in heterologous systems (Randolph V. Lewis 1996; Huemmerich et al. 2004; Teule et al. 2009). Full-length silk proteins are ≥250kD in size and encoded in at least 10kB of DNA, of which only a handful of full gene sequences have been described (Ayoub et al. 2007). The DNA sequences are GC rich (>70%) and have repeats with near exact sequence identity leading to homologous recombination in standard laboratory hosts (Xu and Lewis 1990; Lewis et al. 1996; Fahnestock and Irwin 1997; Gatesy et al. 2001). These properties make silk genes difficult to maintain in laboratory strains of bacteria and troublesome PCR targets. The silk amino acid sequence is also highly repetitive (Supplementary Information) and after translation is naturally un-structured until polymerized into a fiber by mechanical stimulation (Breslauer et al. 2009). This preference for unstructured protein in combination with the high biotechnology value makes silk an ideal model system for probing the biophysical details of the T3SS.

In this paper we construct the first fully synthetic DNA replica of the full-length major ampullate spidroin 1 gene from *Latrodectus hesperus* (LHF1) using only sequence information from public databases (Ayoub et al. 2007). Additional silk fragment sequences encompassing a wide range of silk mechanical properties are created using DNA synthesis. All genes synthesized in this work encode all known amino acid sequence information for each silk. A PCR method is developed to create a library of size variants to probe secretion and expression in *Salmonella* as a function of protein size.

3.2. Materials and Methods

3.2.1. Strains Used

Salmonella typhimurium SL1344 was used for all secretion and fluorescence experiments. All cloning was conducted using TOP10 *E.coli* (Invitrogen Cat#: C664-11).

3.2.2. Growth Media

Three medias were used for cultures: Lysogeny Broth Miller (LB) (Difco Cat#: 244510) Lysogeny Broth Lennox (L) (Difco Cat#:241210) and SPI-1 Inducing Media (IM) (LB, 0.3M NaCl (Tartera and Metcalf 1993)). All plasmid manipulations were done in Lysogeny Broth media. Media was supplemented with 25 μg/mL Kanamycin, 30 μg/mL Choloramphenicol or 100 μg/mL Ampicillin as needed.

3.2.3. DNA Synthesis

Spider silk gene sequences were obtained from the NCBI database. DNA2.0 used the amino acid sequence to backtranslate an optimized DNA sequence for expression in eubacteria using GeneDesigner software and proprietary algorithms. Fully sequenced synthetic DNA constructs were used for the work presented here and the DNA and amino acid sequences can be found in the Supplementary Information.

3.2.4. PCR of silk genes

Silk gene truncations were generated by PCR amplifications of fragments from the synthesized silk gene DNA. Fragments were designed to increase in size based upon the block structure of the silk (Ayoub et al. 2007) and the starting point was the 3' end of the construct to match other silk genes used in this study. This was performed by using the KOD polymerase (EMD4Biosciences Cat#71975) according to the manufacturers instructions to produce PCR products as shown in Figure 5B. Fragments were purified by agarose gel extraction using the Zymo DNA Gel Clean extraction kit (Zymo Research Cat#4001) and cloned into the pCASP vector between the HindIII and XbaI sites (Widmaier et al. 2009).

3.2.5. Western Blots

The proteins to be detected were engineered to contain a C-terminal 3xFLAG epitope tag for detection. Samples were prepared in SDS loading buffer under reducing conditions and run on 10% or 12% polyacrylimide gels (Lonza Biosciences). The gels were transferred to PVDF membranes (BioRad Cat#: 162-0177) and blocked for 1-8 hours in TBST-1% Bovine Serum Albumin (Sigma Cat#: A3294-50G) (TBS-Tween20) with shaking. The anti-FLAG antibody (Sigma Cat# F3165) was used as the primary antibody. The anti-FLAG antibody was mixed 1:7,500 in TBST and allowed to bind for 1 hour. After each antibody step, membranes were rinsed three times with deionized water followed by three 10-minute washes in TBST. The secondary antibody (Sheep anti-mouse HRP [Jackson ImmunoResearch Cat#: 515-035-003]) was added in TBST at 1:5,000 and allowed to bind for one hour. Development was done using an enhanced

chemiluminescent substrate for HRP (Pierce Cat#: 32209) and captured on film (Kodak: Cat#:178-8207).

3.2.6. Quantitative Western Blots

Quantitative western blots were run using an identical protocol as above until the secondary antibody step. The one critical change is using a Hybond low fluorescence PVDF membrane (Amersham Biosciences Cat#: RPN303LFP). A Cy-5 conjugated secondary antibody (GE Healthsciences Cat# PA45010) was diluted 1:5,000 in TBST and processed as above. Final washed blots were dried on 8x8 inch glass plates for 2 hours or complete dryness. Development was done using a Typhoon 9400 variable mode imager to generate gel images (Gingrich et al. 2000; Biosciences 2001). Gel densitometry was done using ImageJ 1.41 (NIH). Each sample was run with duplicate lanes loaded with 10µL of 1x ADF2 supernatant or 0.1x ADF2 lysate (Widmaier et al. 2009) to enable an internal standard of signal across gels. Protein yields were determined using the average of these two bands was used to determine a single point quantitation standard based upon the previously published concentrations of these samples. An exception was made for the full-length LHF1 Lysate gel, Supplementary Information, and a single standard band was run due to limitations of material. Raw gel data for all quantitative westerns in this study are provided in the Supplementary Information.

3.2.7. Secretion Assay

The SPI-1 T3SS is induced in high-salt LB (0.3M NaCl) and uninduced in LB-Lennox (L broth, 0.085M NaCl) (Tartera and Metcalf 1993). Cells were plated on L-broth agar plates from frozen stock and grown overnight. Single colonies were picked and grown 13 hours overnight in 5 ml liquid L broth. The overnights were diluted 1:100 (50μL) into 5mL fresh L-broth and grown for 150 minutes at 37°C and 250 rpm. The cultures were diluted a second time 1:10 into 50mL of inducing media in a non-baffled 250mL glass flask supplemented with antibiotics and the appropriate IPTG concentration. The data generated in Figures 2, 4, and 5 used IPTG supplementation during the dilution to 50mL, a separate protocol is described for the secretion timecourse. The cultures were grown at 37°C for 6 hours at 160 rpm to correspond to match the timing used for the promoter activity assay. The time point T = 0 is immediately after the second dilution step.

3.2.8. Extracellular Protein Collection

Supernatants were harvested by spinning cultures at 3500g for 30 minutes followed by vacuum filtration through 0.45μm cellulose acetate filter unit (Corning Cat# 430314). The supernatant samples of proteins fused to the SptP N-terminal tag were used for the Western blots without concentration*. The ADF4 silk monomers could not be visualized in 1x supernatant, so these samples were concentrated 30x by centrifugal filter (Amicron Cat# UFC801024) to quantitatively measure the background.

3.2.9. Preparation of Cell Pellet Samples

Cell pellets were collected to measure the amount of protein expressed, but not secreted. 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 0.2 mg/mL lysozyme (MP Biomedicals Cat#100834) and incubation for 15 minutes at room temperature followed by a 30 minute freeze/thaw cycle at -80^{o}C. Finally, the lysate was sonicated for 2 minutes in 1 second pulses with 25% amplitude. The resulting mixture was spun to remove insoluble debris at 3500g for 20 minutes. A 1mL sample of the soluble fraction was collected and frozen for quantitative western blot analysis.

3.2.10. SPI-1 Dependent Expression of psicA vs Time

The hybrid *psicA*02 promoter driving the expression of the sicP chaperone and sptP167 tagged ADF2 was co-transformed with the pREP-4 plasmid constitutively expressing LacI protein into electrocompetent *Salmonella* SL1344. Three independent colonies were selected and secretion assays were prepared as described above. Each of the colonies was seeded into 5 replicate flasks of inducing media. One flask for each colony was supplemented with 1000μM IPTG every 4 hours and harvested 4 hours later for 20 hours total. At each sample point OD600 was measured and secreted protein and cell pellet samples were collected as described (materials and methods).

3.2.11. IPTG Titration of Hybrid psicA Promoters

Two series of IPTG titration experiments were run. Promoter activity versus IPTG titration was run with the *psicA*01 and *psicA*02 promoters using a green fluorescent protein (GFPmut3) reporter and flow cytometry measurements. Samples were grown from 3 independent colonies as described in the secretion assay section with 0, 10, 100, or 1000 μM IPTG. At 2,3,4,5, and 6 hours the OD600 of each culture was measured and a 500μL aliquot of cells were collected. Cells were centrifuged at 5000rpm for 3 minutes at room temperature in an Eppendorf 5415D micro centrifuge. The media was vaccum aspirated from each sample without disturbing the cell pellet. The pellet was resuspended in 300μL of ice cold phosphate buffered saline supplemented with 2mg/mL kanamycin. All samples were stored at 4°C for flow cytometry analysis.

For secretion and expression samples IPTG titration experiments were run using the secretion assay protocol. Supernatant and lysate protein samples were collected as described after 6 hours of growth in 50mL of inducing media. All samples were analyzed by quantitative western blotting as described in materials and methods.

3.2.12. LacI Expression

LacI protein was constitutively expressed on a second plasmid based upon a p15a compatibility class origin and kanamycin resistance marker (partsregistry.org Part#pSB3K3). Three different strength constitutive promoters were tried to titrate the level of LacI protein in the cell. The commercial vector, pREP-4 used for most studies here. In addition the promoters, BioBrick part numbers J23150 (Kelly et al. 2009) and I14032 were cloned in front of the lacI open reading frame using inverse PCR

mutagenesis. After screening by GFP reporter assay pREP-4 was used for most measurements made in this study. However repression and dynamic range of the *psicA*01 promoter significantly improved with the J23150 driven LacI (data not shown).

3.2.13. Construction of LacO psicA Hybrid Promoters

A library of hybrid *psicA* promoters containing one or two lac operator binding sites were constructed using PCR mutagenesis. The *psicA* promoter was cloned from the *Salmonella* genome as previously described (Widmaier et al. 2009). The first O1 lac operator (Muller et al. 1996) was introduced 3' to the -10 hexamer of the promoter (Figure 17). The O1 palindrome was generated by asymmetric oligonucleotide primers designed to incorporate the 20bp operator while deleting 20bp of natural *psicA* DNA. Incorporation was done by inverse PCR using the TaKaRa DNA polymerase kit (Clontech Laboratories Cat#TAKRR002M). Untemplated 3' TA nucleotide overhangs were removed using T4 DNA polymerase (New England Biosciences Cat#M0203S) to generate blunt ends. Linear blunt ended DNA was ligated using T4 DNA ligase (Invitrogen Cat#15224-017) in the rapid ligation reaction according to the manufacturer's protocol. A second O1 operator binding site was introduced with spacings of 71, 82, and 93bp 5' of the first lac operator using the same method (Figure 17). For the 71 and 93bp spacings additional promoter mutants were generated using the described PCR protocol to insert or delete up to 3 bp at single base resolution immediately 5' to the invF:sicA operator binding site.

3.2.14. Flow Cytometry Measurements of Hybrid Promoters

All samples expressing a GFPmut3 fluorescent protein were measured by flow cytometry on an LSR II with HTS attachment in 96 well plate mode. For each sample 100,000 counts were collected and processed using FlowJo v7.2 (Tree Star, Inc). Samples were gated for forward and side scatter and an axis population on the FITC-A channel was gated out. Representative population histograms can be found in the Supplementary Information.

3.2.15. Library Screening of Hybrid psicA Promoters

The full library of hybrid *psicA* promoters at all spacings were screened using the GFP reporter assay. Each promoter was screened in the presence of LacI protein with either 0 or 1000μM IPTG to determine the dynamic range of repression versus spacing (Figure 18).

3.3. Results

The natural *psicA* promoter from the Type III Secretion System (T3SS) in *Salmonella* (SPI-1) was used as a starting point to design and construct a series of hybrid promoters which incorporate sensitivity to lac repressor protein (LacI) (Figure 11C, Figure 17). Initially four promoters were constructed, one variant with a single lac operator site (*psicA*01) and three with dual operator binding sites at 71 base pair (bp), 82bp, and 93 bp separation in accordance with the literature (Muller et al. 1996). These promoters were screened using a green fluorescent protein (GFP) reporter in a SPI-1 activation assay (materials and methods). The *psicA*01, 71bp and 93bp promoters

showed functionality similar to wild-type *psicA* promoters with some variation in the magnitude of activation. The 82bp promoter showed no clear activation and was not pursued further. Additional adjustments were made to 71bp and 93bp by adding or removing 3bp in single bp resolution to determine the sharp repression maxima relative to spacing expected from LacO looping (Muller et al. 1996). Alterations in spacing were made immediately 5' to the invF:sicA operator that is critical for *psicA* signal integration from the SPI-1 network (Figure 17). The entire library of operator spacing mutants was screened for activity and the most suitable member selected for further characterization (Figure 18). The promoter most suitable for subsequent testing was the 68bp spacing given its extremely low signal off state and is hereafter referred to as *psicA*02.

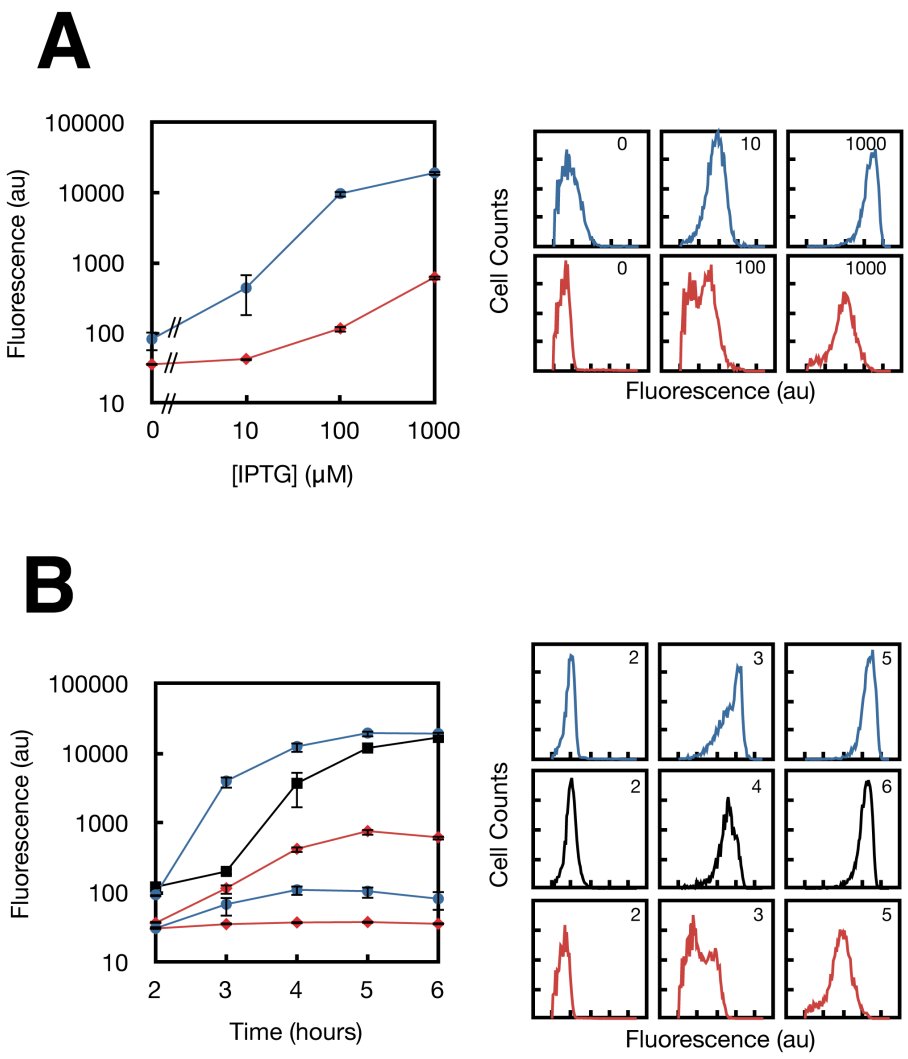


Figure 12

The function of hybrid *psicA* promoters in response to IPTG induction and time. Each point shown is an average of three independent colonies and error bars are +/- one standard deviation. **A.** The response of *psicA*01 (blue circles) and *psicA*02 (red diamonds) over a range of IPTG concentrations are shown. Each promoter was tested by promoter reporter assay (materials and methods) using green fluorescent protein and flow cytometry. Representative histograms are shown to the right representing populations for low, medium, and high induction. The concentration of IPTG inducer added is labeled in units of μM. The x-axis tick marks for these histograms are in $\log_{10}$ scale and start at the axis at 10^0 fluorescence units. **B.** The response of *psicA* (black squares), *psicA*01 (blue circles), and *psicA*02 (red diamonds) with respect to time and inducer are shown. The bottom curves for *psicA*01 and *psicA*02 are with no inducer added while

the top curves are with 1000 μM IPTG. The native *psicA* promoter does not respond to IPTG. The histograms are in the same format as **(A)** with the timepoint for each histogram (in hours) labeled.

Both promoters *psicA*01 and *psicA*02 were characterized for function with regards to the two inputs they integrate, small molecule inducer (IPTG) and functional SPI-1 secretion. Confirmation that the hybrid promoters still functioned with similar dynamics to wild-type *psicA* was done by GFP reporter assay in the absence of LacI repressor (data not shown). A titration curve was performed in the presence of LacI expressed on a plasmid. Increasing concentrations of IPTG added to *psicA*01 and *psicA*02 driving the expression of GFP when grown in a SPI-1 activation assay (Figure 12A). To confirm that the hybrid promoters integrate these two signals appropriately samples were grown with the GFP reporter, LacI repressor protein, and with or without 1000μM IPTG inducer (Figure 12B).

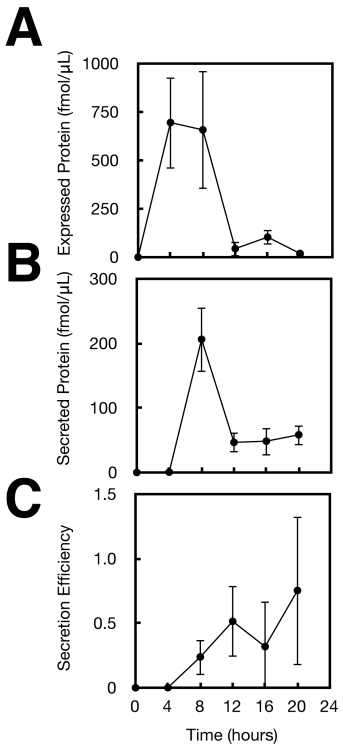


Figure 13

Probing function of Type 3 Secretion over time using the *psicA*02 hybrid promoter. Each point is the average of three independent colonies and error bars represent +/- one standard deviation. **A.** ADF2 silk protein expression measured at 4-hour resolution up to 20 hours. **B.** ADF2 silk protein secretion measured at 4-hour resolution for up to 20 hours. **C.** The secretion efficiency of Type III Secretion, defined as secreted ADF2 protein divided by total ADF2 expressed, is shown over the full 20 hours.

The extremely tight repression (Figure 12B, bottom red diamonds) of the *psicA*02 promoter makes it an ideal system for determining the time dependent profile of SPI-1 secretion. The *psicA*02 promoter was cloned into the SPI-1 secretion plasmid previously described (Figure 11B) driving the secretion of the spider silk ADF2 using the method of Gibson for scarless cloning (Gibson et al. 2009). This reporter was grown in *Salmonella* with inducer added in four-hour intervals (materials and methods) and protein titer was measured using a FLAG epitope tag and quantitative western blotting. The result is a 20-

hour timecourse of SPI-1 dependent expression and secretion of ADF2 protein with four-hour resolution. The expression of protein by *psicA* is at the maxima for the first 8 hours (Figure 13A) and drops off drastically thereafter. Secreted protein is very low to undetectable up to the 4-hour point and at its maximum by 8 hours (Figure 13B). After 8 hours secretion drops off but not completely. A combination of these two measurements gives the secretion efficiency of *Salmonella* versus time (Figure 13C). Secretion efficiency is calculated as the ratio of secreted ADF2 protein to total ADF2 protein expression.

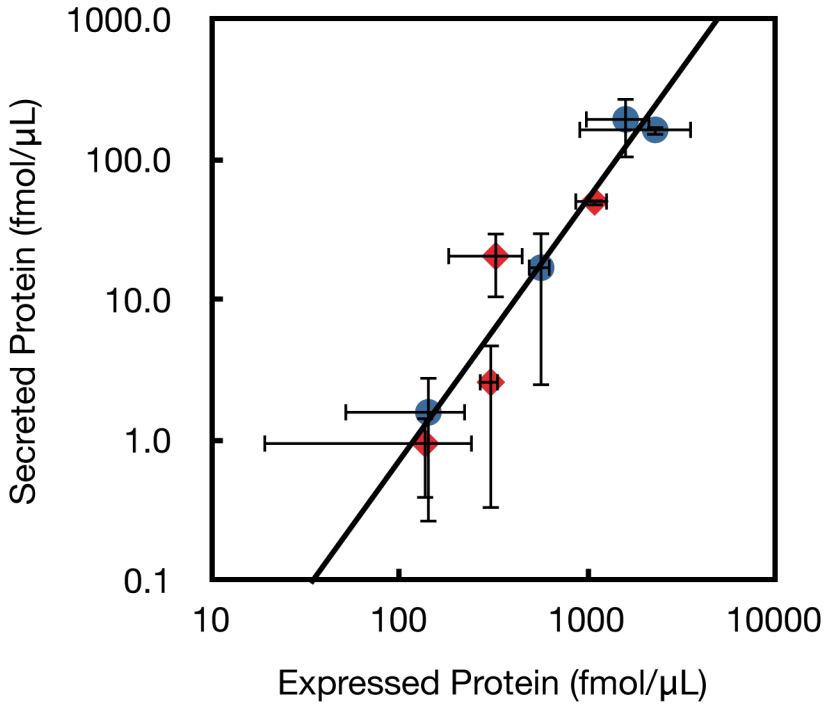


Figure 14

The relationship between protein expression and secretion measured for ADF2 under the control of *psicA*01 (blue circles) and *psicA*02 (red diamonds) is shown. An approximate trendline is overlaid to fit

the data. Each point is the average of three independent colonies and error bars represent +/- one standard deviation.

Using the ADF2 secretion circuit the relationship between protein expression and secretion was measured. A series of inducer (IPTG) titration experiments were performed with *psicA*01 and *psicA*02 to measure ADF2 in the cell lysate and in the culture broth after a 6-hour growth period. A 6-hour growth time was selected to match the GFP characterization experiments presented above. ADF2 concentration was measured using quantitative western blotting as described in the materials and methods. A plot of ADF2 secretion versus expression shows the relationship (Figure 14).

The spider silk proteins were selected as a model system for examining the biophysical limitations of SPI-1 secretion substrates. This system has advantages including naturally unfolded proteins and a modular design for altering protein size. Using sequence information from Genbank the amino acid sequences of several silk genes were used to generate computationally optimized DNA sequences for expression in eubacteria. Each optimized gene was synthesized using chemical gene synthesis and cloned into the pCASP system (Figure 11B). The silk genes synthesized for this study include the full-length gene encoding *Latrodectus hesperus* MaSp1 (LHF1) as well as fragments incorporating all available sequence information for *Nephila clavipes* flagelliform (NCF1), *Araneus diadematus* MaSp2 (ADF4), and *Bombyx mori* heavy chain fibroin (BMF1) (full sequence information in the Supplementary Information). The synthetic genes encode the natural amino acid sequence while reducing repetitiveness in the DNA sequence to prevent homologous recombination.

To further improve the pool of length variants the LHF1 and NCF1 constructs were expanded into libraries of truncation mutants. As identified in the literature LHF1

has four characteristic amino acid block repeat sequences which repeat in a regular order throughout the protein (Ayoub et al. 2007). The full-length synthetic LHF1 construct was copied into fragments using polymerase chain reaction to give products ranging from 40 blocks to 8 blocks in increments of 4 blocks (Figure 15A, Figure 15B). Each construct contains the C-terminal hydrophilic domain (Figure 15A). The LHF1 truncation constructs range from 330 to 1250 amino acids. A similar process was carried out with NCF1 however there are fewer block repeats in this silk type and only two truncation mutants were made and both include the C-terminal domain (Figure 15A).

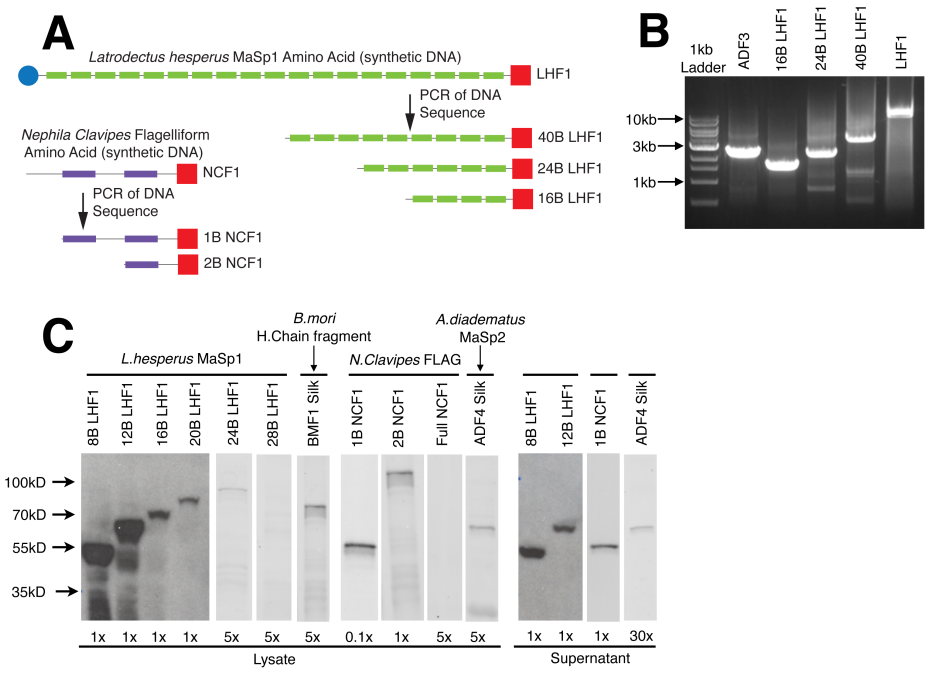


Figure 15

The construction and testing of many size variants for different silk types. Silk DNA sequence information was obtained from Genbank for multiple silk genes including *Latrodectus hesperus* Major Ampullate Spidroin 1 (LHF1), *Nephila clavipes* Flagelliform silk (NCF1), *Araneus diadematus* Major Ampullate Spidroin 2 (ADF4), and *Bombyx mori* Heavy Chain Fibroin (BMF1), all of which were computationally optimized against homologous recombination and synthesized using chemical gene synthesis. **A.** Long genes such as full length LHF1 and the partial clone of NCF1 were further divided

using PCR to create libraries of truncation mutants based on distance from the C-terminus of the amino acid sequence they encode. **B.** A DNA gel showing PCR products for various truncations. Including ADF3 as an external control for a different silk sequence. Representative LHF1 truncations show increasing size of the encoded polypeptide. **C.** Western blots show the detection of silk protein expression and secretion. Each lane is labeled for the concentration of sample loaded and identity.

The entire library of silk genes was cloned into the pCASP vector and a series of secretion assays were run. Samples of supernatant protein and lysate were measured by FLAG epitope tag and quantitative western blot (Figure 15C). The data obtained for secretion and expression are plotted versus number of codons expressed (Figure 16A, Figure 16B). Previously measured silk secretion and expression titers are plotted as well. When secretion efficiency is calculated for this library a trend of lower secretion efficiency for longer proteins is established (Figure 16C).

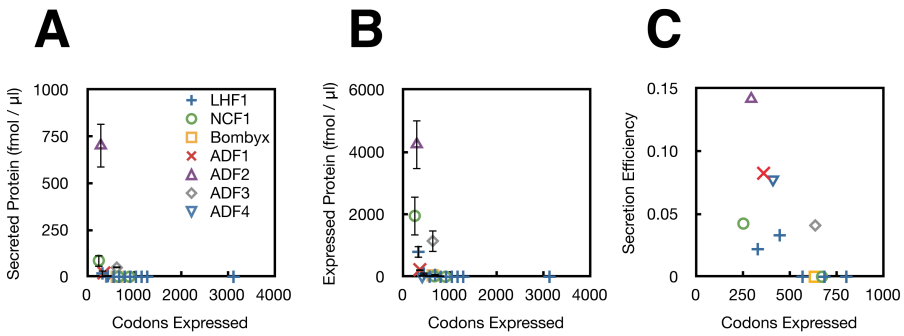


Figure 16

All silk proteins in this study are measured for expression and secretion. Each point is the average of three independent colonies and error bars represent +/- one standard deviation. Additional silk proteins measured for this system in a previous publication are also plotted for reference (Widmaier et al. 2009). **A**. Secretion of each protein is plotted vs the protein size in number of codons. Different spider species are shown by different symbols *Latrodectus hesperus* MaSp1 (blue +), *Nephila clavipes* Flagelliform (green circles), *Bombyx mori* cocoon silk (yellow square), *Araneus diadematus* ADF1 (red X) ADF2 (purple

triangle) ADF3 (grey diamonds) and ADF4 (blue triangle). **B**. Expression of all proteins is shown versus number of codons. **C.** The secretion efficiency, defined as the ratio of [silk secretion titer / total silk expression] for all silks with a detectible expression titer are shown.

3.4. Discussion

In this work the Type III Secretion System in *Salmonella typhimurium* (SPI-1) is improved as a tool for cellular engineering. The ability to connect synthetic control mechanisms with the genetic network of SPI-1 is demonstrated and used to answer outstanding questions in the biology of the secretion system. A number of silk genes are constructed and used as a model system to probe the size dependence of expression and secretion in *Salmonella*. This data provides a basis for further engineering of SPI-1 to enable a range of protein secretion technologies such as therapeutic peptide or biomass degradation protein export.

The ability to interface synthetic control systems with natural functional genetic modules is a key focus of the field of synthetic biology. The ability to create desired information processing motifs within cells is growing rapidly, however the ability to harness complex biological functionalities has grown slowly (Purnick and Weiss 2009). This work introduces a new method for transcriptional control based upon LacO looping for tight repression of target promoters. This method should be generalizable to a large number of target promoters to add a user defined control feature. This DNA looping system can be utilized in multiple configurations where changes to the state of repressor protein can be accomplished by induction, degradation, or titration with small molecule inducer. Each approach to modifying the repressor protein has a different

dynamic effect on the functional state of a target promoter. For example, LacO looping could be used to incorporate various feedback and feedforward mechanisms from cellular sensing networks and translate these into LacI production or degradation, which becomes a functional output (ie: protein secretion). Most critically, using DNA looping to modulate promoter function provides a mechanism to control complex promoters in a cell by adding additional input integration while preserving natural functionality.

The DNA looping approach applied to SPI-1 provides a model system to probe the secretion ability of *Salmonella* in cell culture over time. The off state of the *psicA*02 promoter is indistinguishable from white cells in the GFP activity assay and shows much better baseline control than even the natural *psicA* promoter (Figure 12B). Using this feature the secretion of proteins by SPI-1 is shown to have a very clear maximum at early stationary phase (Figure 13B) and fall off dramatically. Inspection of the trend in the efficiency of protein secretion over time (Figure 13C) shows a general increase in the total percentage of cellular ADF2 secreted as a function of time. Presumably this means that while secretion complexes are able to continue protein export, the *psicA* promoter is unable to continue expression into stationary phase. Several reasons can be speculated here including a repression mechanism in preparation for SPI-2 induction.

This work further refines the role of SPI-1 secretion as a tool for protein export. The model silk protein system is used to define the size limits of expression and secretion in *Salmonella*. A correlation between protein expression and secretion is found (Figure 16) and further testing of expression optimized constructs may provide clues about how to secrete larger proteins more efficiently. Additional data points with more

divergent amino acid sequences would greatly add to the robustness of this characterization.

The field of synthetic biology aspires to reliably engineer biological systems while utilizing powerful functionalities evolved by nature within synthetic systems. The key feature of future systems is the exchange of information between endogenous regulatory circuits and the synthetic genetic circuits used to hijack cellular functionality. Here we describe one such method using LacO looping and apply it to the study of Type III Secretion dynamics over time.

3.5. Supplemental Information for Chapter 3

3.5.1. Construction of Hybrid psicA Promoters

The natural *psicA* promoter (Figure 17) was cloned by PCR from *Salmonella typhimurium* SL1344 genomic DNA and is the basis for LacO looping engineering. The lac operator DNA sequence was introduced by site directed mutagenesis (materials and methods) and replaced 20bp of DNA from the wild-type *psicA* promoter. The first variant constructed was *psicA*01 which was further modified to create 71bp, 82bp and 93bp variants based upon optimal spacing found in (Muller et al. 1996). The optimal spacing found in this publication was 70.5bp, 81.5bp and 92.5 bp, so a series of single bp spacing mutants were constructed by site directed mutagenesis. Additional bases (up to 3bp) were added or deleted immediately 5' to the invF:sicA operator site (Figure 17). Interestingly a different optimum spacing for the *psicA* promoter was found compared to Müller et al (Figure 18).

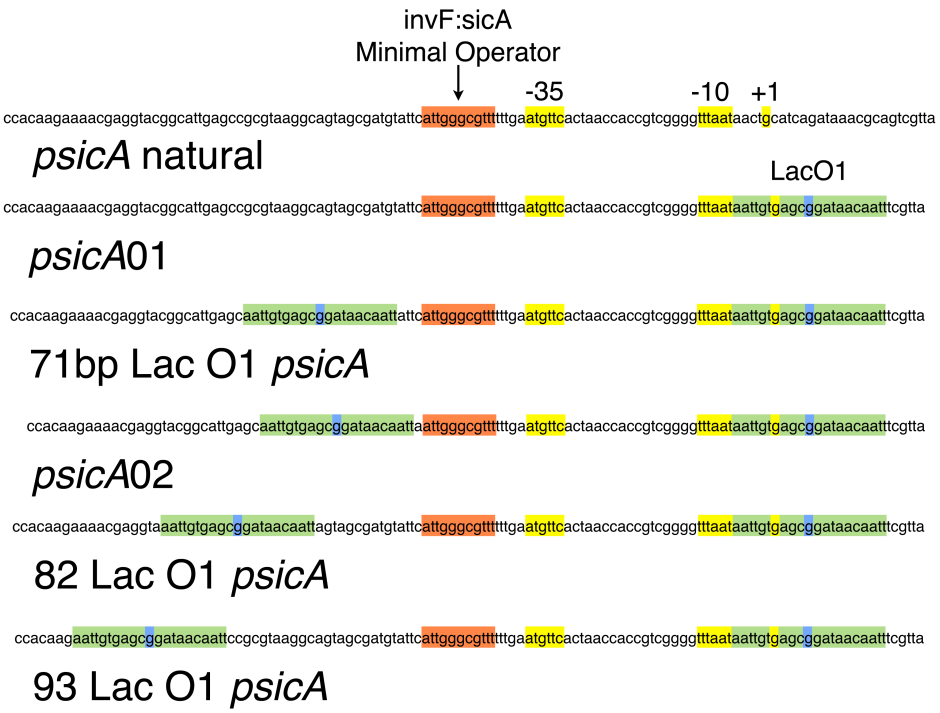


Figure 17

The sequence of the natural *psicA* promoter with critical features highlighted; +1, -10, -35 nucleotides (yellow boxes), invF:sicA operator binding site (orange box) (Darwin and Miller 2001). Additional colored boxes (green with blue base) show the placement of the O1 operator DNA binding sites in the sequence to obtain additional system control. Various lac operator spacings were constructed as measured from center to center (blue guanine nucleotide) and are consistent with the literature for enhanced promoter control (Muller et al. 1996). Small changes in spacing were achieved by adding or deleting single bases immediately 5' of the invF:sicA minimal binding box. An example of the spacing changes is shown by comparing *psicA*02, a three base deletion, with 71bp LacO *psicA*.

3.5.2. Measurement of All Hybrid Promoter Function

The library of LacO spacing mutants were assayed for activity by expressing green fluorescent protein (GFP) in *Salmonella* and measured using flow cytometry (materials and methods). The library was measured twice, once with no IPTG in the culture media and a second time with 1000μM IPTG. The mean fluorescence data is

shown in Figure 18A with and without IPTG. The ratio of induced (+ IPTG) and uninduced (-IPTG) is shown in Figure 18B. The sharp repression maxima as a function of operator spacing resemble the pattern found by Müller et al and are an indicator of DNA looping.

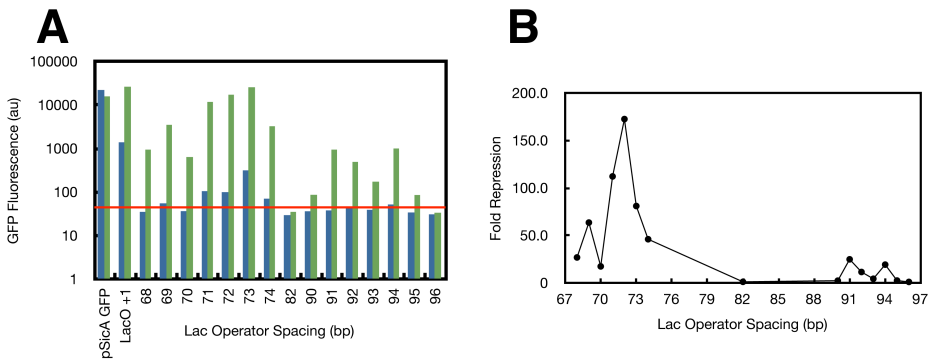


Figure 18

A measure of dynamic range for a series of *psicA* promoter hybrids. Numerous spacing variants of the *psicA* promoter were created (Figure 17) and tested in the presence or absence of 1000μM IPTG. **A.** All promoters in this study are measured with (green bars) and without IPTG (blue bars) using a GFP reporter and flow cytometry measurements (materials and methods). The mean fluorescence of 100,000 events is shown here. The red line is the average fluorescent value for non-GFP expressing cells (white cells) in the assay. **B.** A plot of the ratio of +/- IPTG vs lac operator spacing.

3.5.3. Logic Function of Hybrid psicA

The *psicA*02 promoter was examined using the GFP reporter assay to determine the Boolean logic function. The two inputs, IPTG and LacI were varied and the output shown in Figure 19.

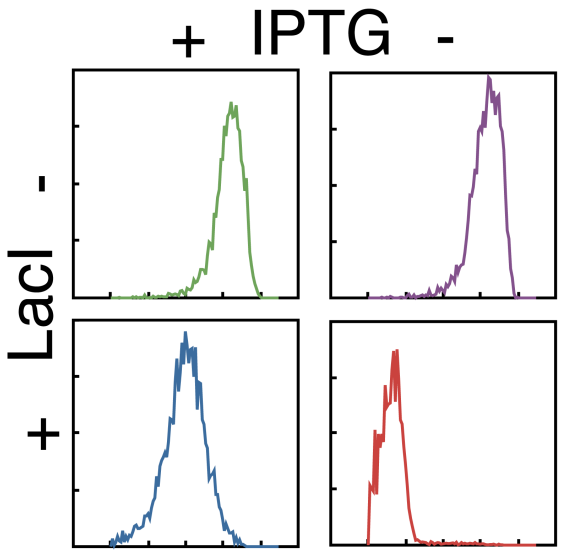


Figure 19

The hybrid *psicA* promoters behave as an NOT AND logical gate. Logic for the *psicA*02 promoter is shown. In the absence of LacI protein expression the promoter is strongly on with or without IPTG (green and purple). When expressing LacI tetramers in the presence of 1000µM IPTG the promoter expresses GFP ~20 fold (blue) over the 0µM IPTG state (red). In the presence of LacI tetramers and the absence of IPTG the GFP output of the promoter is indistinguishable from white cell fluorescence. All graphs have y-axis units of number of cells and x-axis units of fluorescence (au) in $\log_{10}$ scale with 10^0 on y-axis.

3.5.4. SPI-1 Secretion vs Time

Prior to testing the secretion of ADF2 over a 20 hour timecourse (Figure 13) a preliminary test was made to ensure 4 hour windows of expression would induce the *psicA*02 promoter appropriately. The hybrid psicA02 promoter driving the expression of GFP was co-transformed with the pREP-4 plasmid constitutively expressing LacI protein into electrocompetent Salmonella SL1344. The cells were prepared as in a secretion assay (materials and methods). At T=0 the inducing media contained 0µM IPTG. At 4 hours 500µL aliquot of cells were collected (Figure 20A) and then 1000µM IPTG was

added. Cells were centrifuged at 5000rpm for 3 minutes at room temperature in an Eppendorf 5415D micro centrifuge. The media was vacuum aspirated from the sample without disturbing the cell pellet. The pellet was resuspended in 300µL of ice cold phosphate buffered saline supplemented with 2mg/mL kanamycin. All samples were stored at 4°C for analysis. At 8 hours second 500 µL sample was collected and prepared in an identical method before flow cytometry analysis (materials and methods) (Figure 20B).

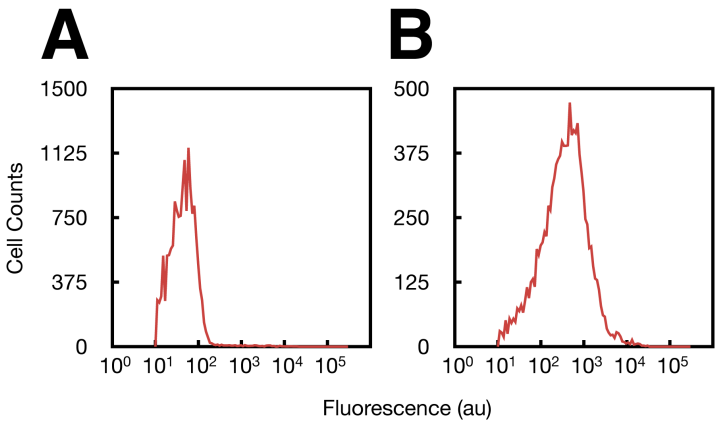


Figure 20

Histograms showing the repression and induction of the *psicA*02 promoter vs time for a GFP promoter assay. **A.** A culture of *Salmonella* with the *psicA*02 plasmid grown for 4 hours with LacI and without IPTG. **B.** 1000µM IPTG was added and the GFP fluorescence of the population at 8 hours is shown.

3.5.5. Secretion and Expression Titer Quantitation

The data to generate Figure 14 is a combination of the ADF2 expression and secretion data shown in Figure 21. This data was collected by densitometry analysis of quantitative western blots.

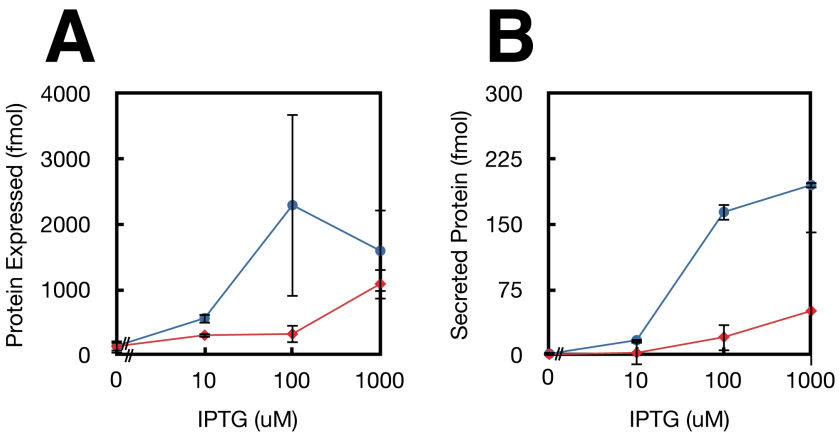


Figure 21

Expression **(A)** and secretion **(B)** of ADF2 protein under the control of the *psicA*01 (blue circles) and *psicA*02 (red diamonds) promoters in the presence of various concentrations of inducer (IPTG). Protein titers are measured by quantitative western blotting (materials and methods). Individual gels this data was process from are found in Figure 22.

3.5.6. Quantitative Western Blots Used in This Study

All quantitative western blots used in this study are copied below. Images are the result of scans for the Cy-5 dye on a Typhoon variable mode imager. The photomultiplier tube sensitivity was adjusted to prevent detector oversaturation for each image. The figures that were generated densitometry data from the presented gels are listed in the captions Figure 22 - Figure 24Error! Reference source not found..

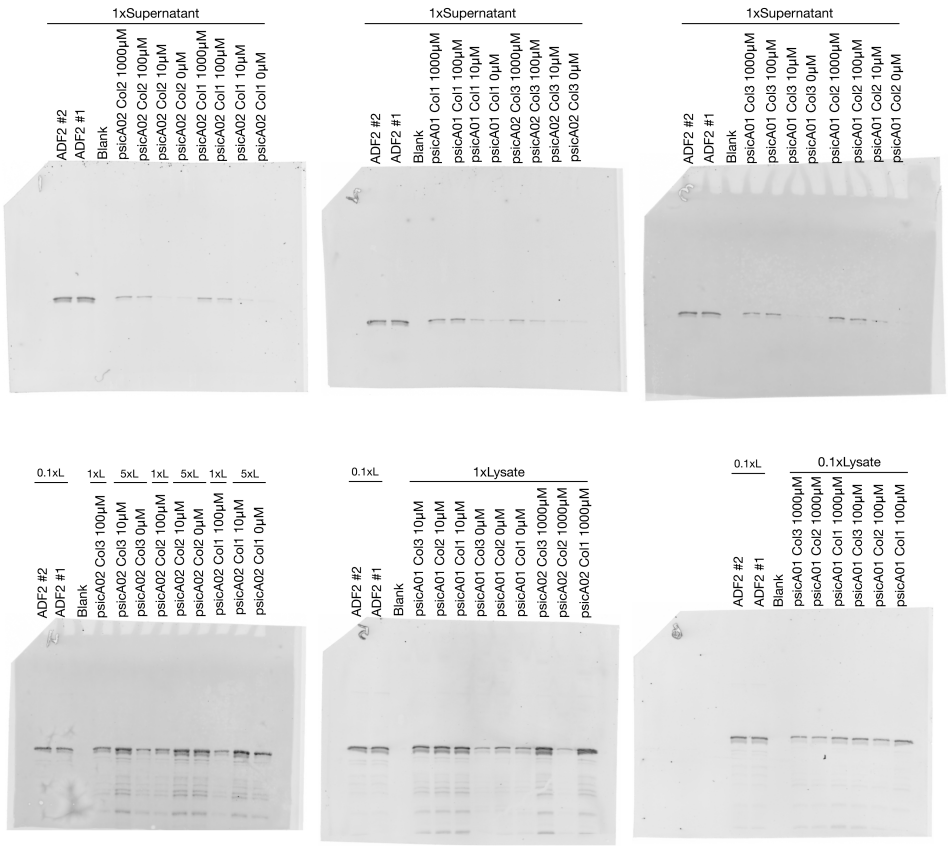


Figure 22

All gels used for quantitative western blotting analysis to determine the curves in Figure 14 and Figure 21. Each blot is labeled for the internal control bands (ADF2), the sample band with colony number and IPTG inducer concentration. The concentration of sample loaded is labeled above the lanes. Each gel has a 1x ADF2 supernatant or 0.1x ADF2 Lysate sample from to act as an internal standard (Widmaier et al. 2009). Gels were processed according to materials and methods to determine signal intensity.

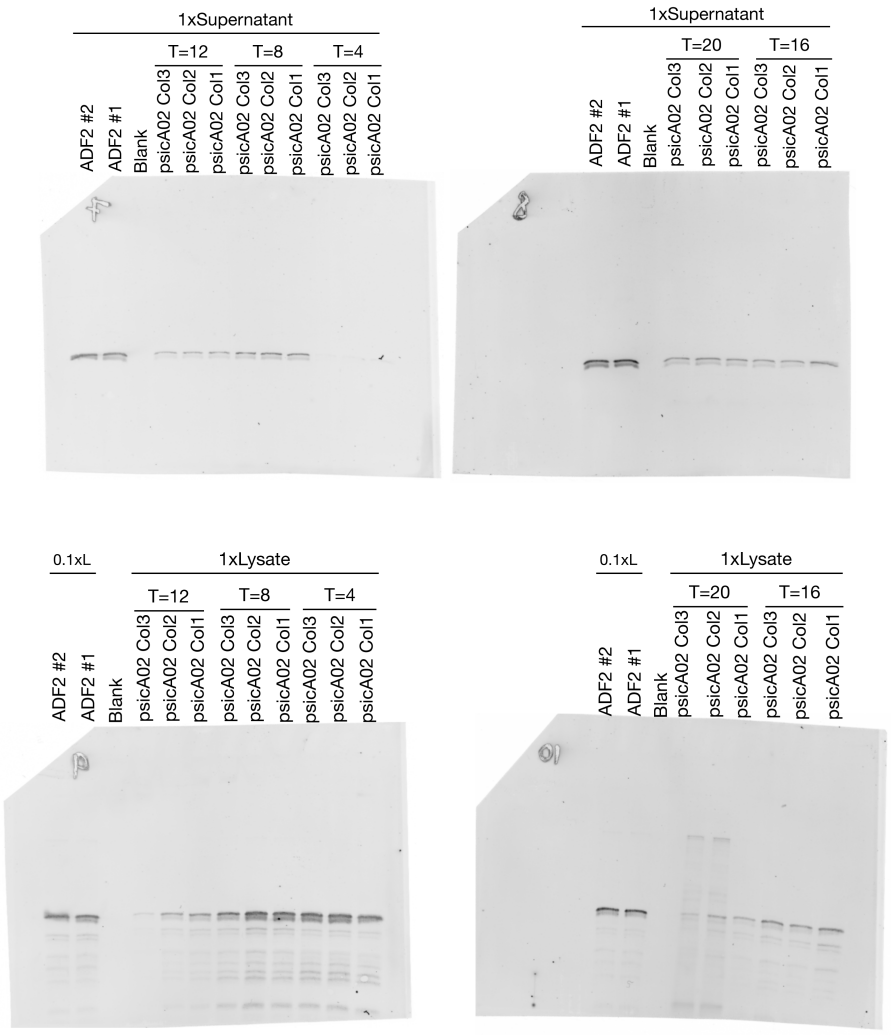
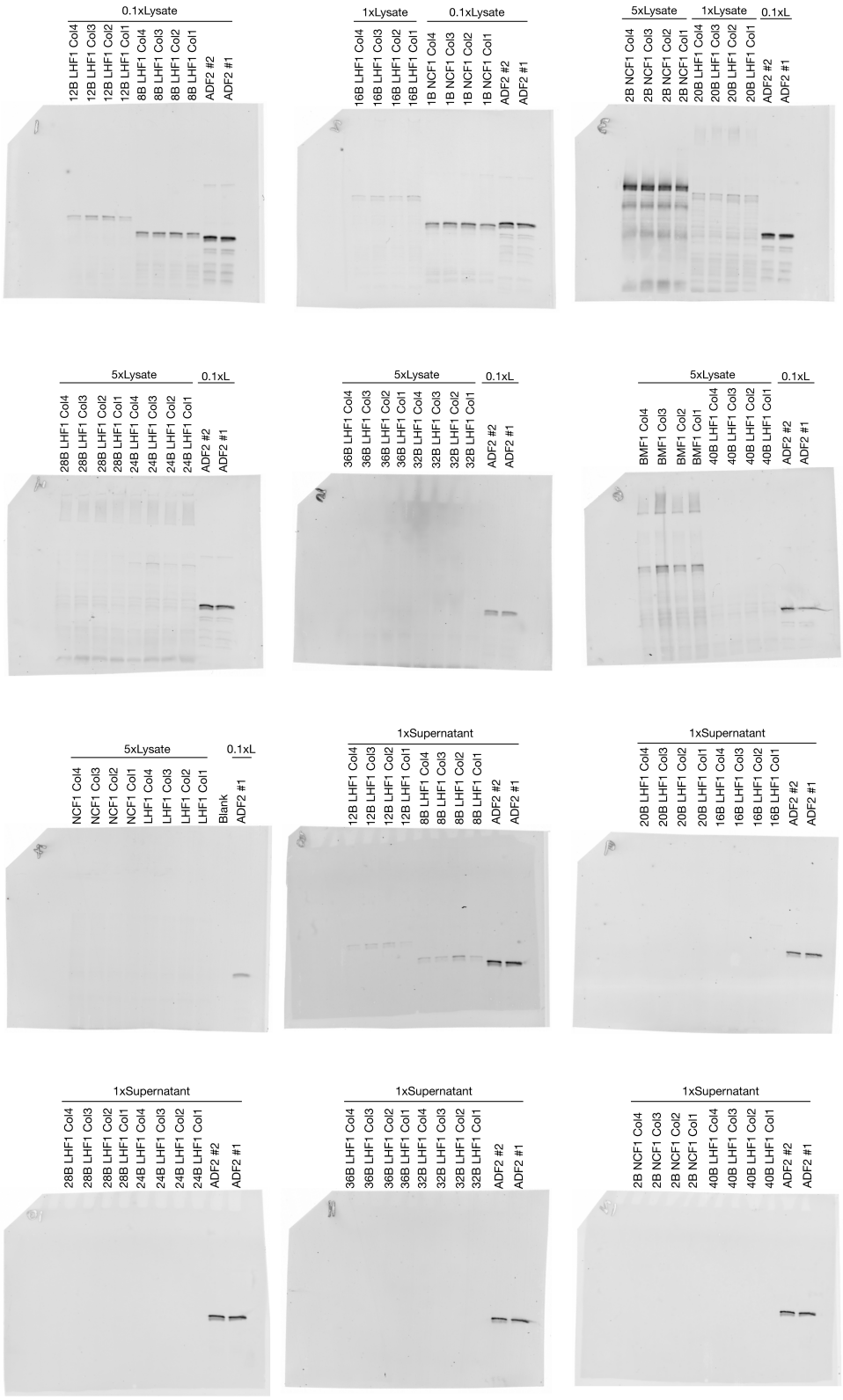


Figure 23

The blots to generate Figure 13 are shown. Each blot is labeled for the internal control bands (ADF2), the sample band with colony number and timepoint of harvest (in hours). The concentration of sample loaded is labeled above the lanes. Each gel has a 1x ADF2 supernatant or 0.1x ADF2 Lysate sample from to act as an internal standard (Widmaier et al. 2009). Gels were processed according to materials and methods to determine signal intensity.


0.1xLysate
12B LHF1 Col4
12B LHF1 Col3
12B LHF1 Col2
12B LHF1 Col1
8B LHF1 Col4
8B LHF1 Col3
8B LHF1 Col2
8B LHF1 Col1
ADF2 #2
ADF2 #1
1xLysate
0.1xLysate
16B LHF1 Col4
16B LHF1 Col3
16B LHF1 Col2
16B LHF1 Col1
1B NCF1 Col4
1B NCF1 Col3
1B NCF1 Col2
1B NCF1 Col1
ADF2 #2
ADF2 #1
5xLysate
1xLysate
0.1xL
2B NCF1 Col4
2B NCF1 Col3
2B NCF1 Col2
2B NCF1 Col1
20B LHF1 Col4
20B LHF1 Col3
20B LHF1 Col2
20B LHF1 Col1
ADF2 #2
ADF2 #1
5xLysate
0.1xL
28B LHF1 Col4
28B LHF1 Col3
28B LHF1 Col2
28B LHF1 Col1
24B LHF1 Col4
24B LHF1 Col3
24B LHF1 Col2
24B LHF1 Col1
ADF2 #2
ADF2 #1
5xLysate
0.1xL
36B LHF1 Col4
36B LHF1 Col3
36B LHF1 Col2
36B LHF1 Col1
32B LHF1 Col4
32B LHF1 Col3
32B LHF1 Col2
32B LHF1 Col1
ADF2 #2
ADF2 #1
5xLysate
0.1xL
BMF1 Col4
BMF1 Col3
BMF1 Col2
BMF1 Col1
40B LHF1 Col4
40B LHF1 Col3
40B LHF1 Col2
40B LHF1 Col1
ADF2 #2
ADF2 #1
5xLysate
0.1xL
NCF1 Col4
NCF1 Col3
NCF1 Col2
NCF1 Col1
LHF1 Col4
LHF1 Col3
LHF1 Col2
LHF1 Col1
Blank
ADF2 #1
1xSupernatant
12B LHF1 Col4
12B LHF1 Col3
12B LHF1 Col2
12B LHF1 Col1
8B LHF1 Col4
8B LHF1 Col3
8B LHF1 Col2
8B LHF1 Col1
ADF2 #2
ADF2 #1
1xSupernatant
20B LHF1 Col4
20B LHF1 Col3
20B LHF1 Col2
20B LHF1 Col1
16B LHF1 Col4
16B LHF1 Col3
16B LHF1 Col2
16B LHF1 Col1
ADF2 #2
ADF2 #1
1xSupernatant
28B LHF1 Col4
28B LHF1 Col3
28B LHF1 Col2
28B LHF1 Col1
24B LHF1 Col4
24B LHF1 Col3
24B LHF1 Col2
24B LHF1 Col1
ADF2 #2
ADF2 #1
1xSupernatant
36B LHF1 Col4
36B LHF1 Col3
36B LHF1 Col2
36B LHF1 Col1
32B LHF1 Col4
32B LHF1 Col3
32B LHF1 Col2
32B LHF1 Col1
ADF2 #2
ADF2 #1
1xSupernatant
2B NCF1 Col4
2B NCF1 Col3
2B NCF1 Col2
2B NCF1 Col1
40B LHF1 Col4
40B LHF1 Col3
40B LHF1 Col2
40B LHF1 Col1
ADF2 #2
ADF2 #1

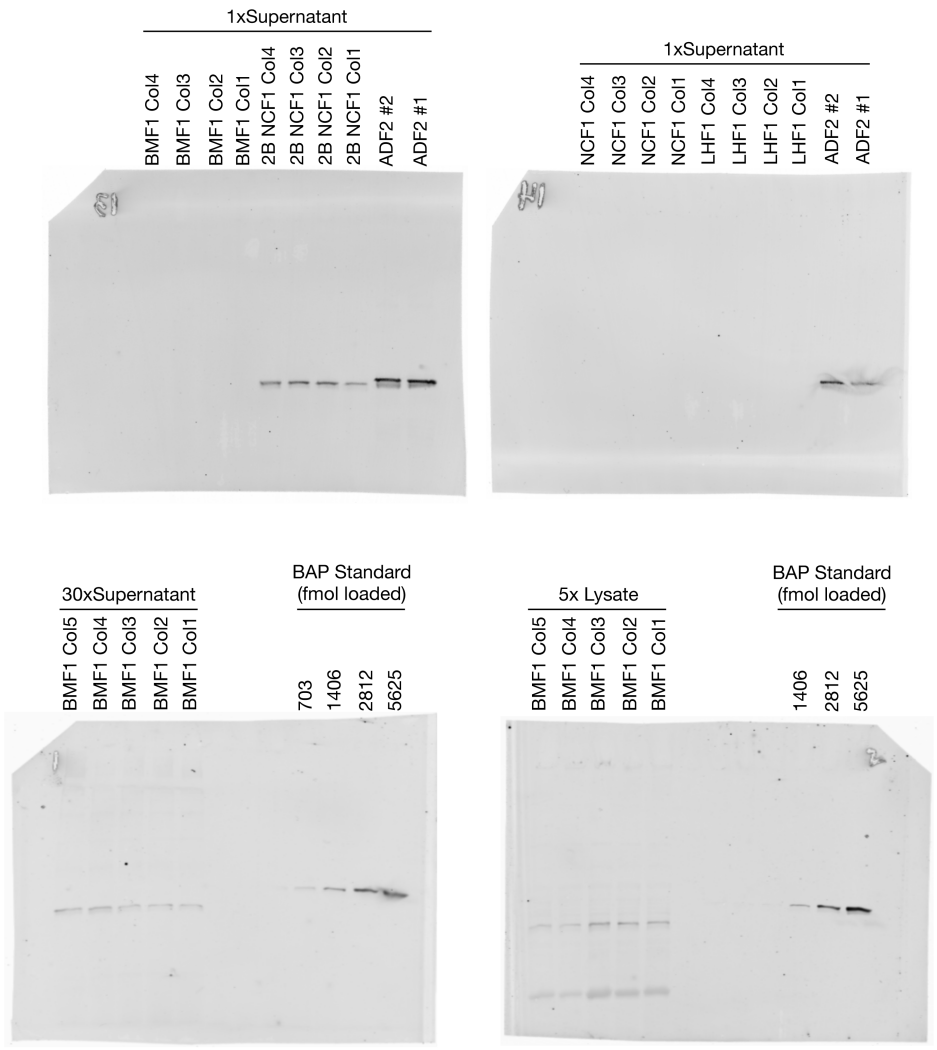


Figure 24

The blots used to generate Figure 15C and Figure 16 are shown. Each blot is labeled for the internal control bands (ADF2), the sample band with colony number and timepoint of harvest (in hours). The concentration of sample loaded is labeled above the lanes. Each gel has a 1x ADF2 supernatant or 0.1x ADF2 Lysate sample from to act as an internal standard (Widmaier et al. 2009). Gels were processed according to materials and methods to determine signal intensity. The blots for ADF4 were run in the previously published method with a external bacterial alkaline phosphatase standard. Calculations were performed as published for the ADF4 blots (Widmaier et al. 2009).

Chapter 4
Conclusion and Future Directions

4.1. Summary of the Findings Presented

In this dissertation the ability to engineer the Type III Secretion System (T3SS) at *Salmonella* Pathogeneity Island 1 (SPI-1) is examined. An interesting circuit within SPI-1 based upon the *psicA* promoter was identified from the literature. Quantitative measurements are made of the dynamics of *psicA* activation, which drive the design of a recombinant protein secretion circuit. This circuit is used to express and secrete spider silk protein, a high value material and a long-standing goal in biotechnology.

The spider silk genes are generated from DNA sequence information only. The constructs presented here are computationally optimized sequences to express protein in *Salmonella* and to prevent homologous recombination of the gene. The optimized genes were chemically synthesized using DNA synthesis and assembly techniques. The SPI-1 T3SS is harnessed to prevent the spider silk protein from forming fibrils within the confined volume of the cell. Expression and secretion of spider silk occurs only when secretion is active.

The ability to add additional control examine the constraints to the secretion circuit is demonstrated by integrating a tunable small molecule signal. Control is added by utilizing LacO looping to tightly control the *psicA* promoter. This control method is shown to work in conjunction with the natural function of *psicA* and creates a Boolean AND gate by integrating the signals from SPI-1 regulation and small molecule input (IPTG). Additional spider silk genes are constructed and used to determine the constraints of SPI-1 secretion. An upper limit of ~800 amino acids for expression and secretion of silk proteins is established.

4.2. Future Directions

Several improvements to this secretion system are clear. First and foremost for usefulness in a biotechnology role is increasing protein secretion capacity per cell. The most promising method to attempt this is an over-expression of needle complexes. One particularly intriguing future direction is replacing the *psicA* system with an entirely synthetic circuit. Many designs might be conceived for utilizing the T3SS as a crucial component in a negative feedback circuit. A secretion tagged repressor molecule would act upon it's own expression and that of a spider silk protein. In this way secretion of the repressor outside of the cell would segregate it from the DNA it acts upon. This negative feedback mechanism would accomplish a similar dynamics as *psicA* without the need for the SPI-1 transcription factors invF and sicA. In many ways it could improve the system to dynamically adapt the expression of silk protein to the secretion capacity of the cell. Hopefully a future researcher can further examine these speculations for improving SPI-1, and fulfill the development of a robust and programmable secretion system.

References

Akeda, Y. and J. E. Galan (2005). "Chaperone release and unfolding of substrates in type III secretion." Nature **437**(7060): 911-5.

Altier, C. (2005). "Genetic and environmental control of Salmonella invasion." Journal of Microbiology **43**: 85-92.

Arcidiacono, S., C. Mello, et al. (1998). "Purification and characterization of recombinant spider silk expressed in Escherichia coli." Appl Microbiol Biotechnol **49**(1): 31-8.

Ayoub, N. A., J. E. Garb, et al. (2007). "Blueprint for a high-performance biomaterial: full-length spider dragline silk genes." PLoS One **2**(6): e514.

Bajaj, V., R. L. Lucas, et al. (1996). "Co-ordinate regulation of Salmonella typhimurium invasion genes by environmental and regulatory factors is mediated by control of hilA expression." Mol Microbiol **22**(4): 703-14.

Bayer, T. S., D. M. Widmaier, et al. (2009). "Synthesis of methyl halides from biomass using engineered microbes." J Am Chem Soc **131**(18): 6508-15.

Bendtsen, J. D., H. Nielsen, et al. (2005). "Prediction of twin-arginine signal peptides." Bmc Bioinformatics **6**: -.

Berks, B. C. (1996). "A common export pathway for proteins binding complex redox cofactors?" Molecular Microbiology **22**(3): 393-404.

Besse, M., B. von Wilcken-Bergmann, et al. (1986). "Synthetic lac operator mediates repression through lac repressor when introduced upstream and downstream from lac promoter." Embo J **5**(6): 1377-81.

Biosciences, A. (2001) "Application Note: Fluorescent Western Blotting." **Volume**, DOI:

Boyd, A. P., N. Grosdent, et al. (2000). "Yersinia enterocolitica can deliver Yop proteins into a wide range of cell types: Development of a delivery system for heterologous proteins." European Journal of Cell Biology **79**(10): 659-671.

Breslauer, D. N., L. P. Lee, et al. (2009). "Simulation of flow in the silk gland." Biomacromolecules **10**(1): 49-57.

Bronstein, P. A., E. A. Miao, et al. (2000). "InvB is a type III secretion chaperone specific for SspA." J Bacteriol **182**(23): 6638-44.

Chen, L. M., G. Briones, et al. (2006). "Optimization of the delivery of heterologous proteins by the Salmonella enterica serovar Typhimurium type III secretion system for vaccine development." Infect Immun **74**(10): 5826-33.

Choi, J. H. and S. Y. Lee (2004). "Secretory and extracellular production of recombinant proteins using Escherichia coli." Appl Microbiol Biotechnol **64**(5): 625-35.

Collazo, C. M. and J. E. Galan (1996). "Requirement for exported proteins in secretion through the invasion-associated type III system of Salmonella typhimurium." Infect Immun **64**(9): 3524-31.

Darwin, K. H. and V. L. Miller (1999). "InvF is required for expression of genes encoding proteins secreted by the SPI1 type III secretion apparatus in Salmonella typhimurium." J Bacteriol **181**(16): 4949-54.

Darwin, K. H. and V. L. Miller (2000). "The putative invasion protein chaperone SicA acts together with InvF to activate the expression of Salmonella typhimurium virulence genes." Mol Microbiol **35**(4): 949-60.

Darwin, K. H. and V. L. Miller (2001). "Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in Salmonella typhimurium." Embo J **20**(8): 1850-62.

Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." Proc Natl Acad Sci U S A **97**(12): 6640-5.

DeLisa, M. P., D. Tullman, et al. (2003). "Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway." Proceedings of the National Academy of Sciences of the United States of America **100**(10): 6115-6120.

Economou, A. and W. Wickner (1994). "Seca Promotes Preprotein Translocation by Undergoing Atp-Driven Cycles of Membrane Insertion and Deinsertion." Cell **78**(5): 835-843.

Ehrbar, K., A. Friebel, et al. (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**(23): 6950-67.

Eichelberg, K. and J. E. Galan (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**(8): 4099-105.

Ellermeier, C. D., J. R. Ellermeier, et al. (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**(3): 691-705.

Enninga, J., J. Mounier, et al. (2005). "Secretion of type III effectors into host cells in real time." Nat Methods **2**(12): 959-65.

Fahnestock, S. R. and S. L. Irwin (1997). "Synthetic spider dragline silk proteins and their production in Escherichia coli." Appl Microbiol Biotechnol **47**(1): 23-32.

Feldman, M. F., S. Muller, et al. (2002). "SycE allows secretion of YopE-DHFR hybrids by the Yersinia enterocolitica type III Ysc system." Mol Microbiol **46**(4): 1183-97.

Fu, Y. and J. E. Galan (1998). "Identification of a specific chaperone for SptP, a substrate of the centisome 63 type III secretion system of Salmonella typhimurium." J Bacteriol **180**(13): 3393-9.

Fu, Z., S. B. Hamid, et al. (2003). "Secretory expression in Escherichia coli and single-step purification of a heat-stable alkaline protease." Protein Expr Purif **28**(1): 63-8.

Galan, J. E. and A. Collmer (1999). "Type III secretion machines: bacterial devices for protein delivery into host cells." Science **284**(5418): 1322-8.

Gatesy, J., C. Hayashi, et al. (2001). "Extreme diversity, conservation, and convergence of spider silk fibroin sequences." Science **291**(5513): 2603-5.

Georgiou, G. and L. Segatori (2005). "Preparative expression of secreted proteins in bacteria: status report and future prospects." Curr Opin Biotechnol **16**(5): 538-45.

Gerchman, Y. and R. Weiss (2004). "Teaching bacteria a new language." Proc Natl Acad Sci U S A **101**(8): 2221-2.

Gibson, D. G., L. Young, et al. (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." Nat Methods **6**(5): 343-5.

Gingrich, J. C., D. R. Davis, et al. (2000). "Multiplex detection and quantitation of proteins on western blots using fluorescent probes." Biotechniques **29**(3): 636-42.

Gosline, J. M., P. A. Guerette, et al. (1999). "The mechanical design of spider silks: from fibroin sequence to mechanical function." J Exp Biol **202**(Pt 23): 3295-303.

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

Gustafsson, C., S. Govindarajan, et al. (2004). "Codon bias and heterologous protein expression." Trends in Biotechnology **22**(7): 346-353.

Guzman, L. M., D. Belin, et al. (1995). "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter." J Bacteriol **177**(14): 4121-30.

Harvey, B. R., G. Georgiou, et al. (2004). "Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries." Proceedings of the National Academy of Sciences of the United States of America **101**(25): 9193-9198.

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. Ingrahamet al. Washington, D.C, ASM press. **2:** 2047-2066.

Hernan, R., K. Heuermann, et al. (2000). "Multiple epitope tagging of expressed proteins for enhanced detection." Biotechniques **28**(4): 789-93.

Higashide, W. and D. Zhou (2006). "The first 45 amino acids of SopA are necessary for InvB binding and SPI-1 secretion." J Bacteriol **188**(7): 2411-20.

Hofmann, S., H. Hagenmuller, et al. (2007). "Control of in vitro tissue-engineered bone-like structures using human mesenchymal stem cells and porous silk scaffolds." Biomaterials **28**(6): 1152-62.

Hong, K. H. and V. L. Miller (1998). "Identification of a novel Salmonella invasion locus homologous to Shigella ipgDE." J Bacteriol **180**(7): 1793-802.

Horton, R. M. (1995). "Pcr-Mediated Recombination and Mutagenesis." Molecular Biotechnology **3**(2): 93-99.

Huemmerich, D., C. W. Helsen, et al. (2004). "Primary structure elements of spider dragline silks and their contribution to protein solubility." Biochemistry **43**(42): 13604-12.

Huemmerich, D., T. Scheibel, et al. (2004). "Novel assembly properties of recombinant spider dragline silk proteins." Curr Biol **14**(22): 2070-4.

Hussain, N. K., S. Jenna, et al. (2001). "Endocytic protein intersectin-l regulates actin assembly via Cdc42 and N-WASP." Nat Cell Biol **3**(10): 927-32.

Jacob, F. and J. Monod (1961). "Genetic regulatory mechanisms in the synthesis of proteins." J Mol Biol **3**: 318-56.

Kalir, S., J. McClure, et al. (2001). "Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria." Science **292**(5524): 2080-3.

Kaniga, K., S. Tucker, et al. (1995). "Homologs of the Shigella IpaB and IpaC invasins are required for Salmonella typhimurium entry into cultured epithelial cells." J Bacteriol **177**(14): 3965-71.

Kapust, R. B., J. Tozser, et al. (2002). "The P1' specificity of tobacco etch virus protease." Biochem Biophys Res Commun **294**(5): 949-55.

Karavolos, M. H., A. J. Roe, et al. (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**(5): 1559-67.

Kelly, J. R., A. J. Rubin, et al. (2009). "Measuring the activity of BioBrick promoters using an in vivo reference standard." Journal of Biological Engineering **3**(4).

Kluge, J. A., U. Rabotyagova, et al. (2008). "Spider silks and their applications." Trends in Biotechnology **26**(5): 244-251.

Knodler, L. A., M. Bertero, et al. (2006). "Structure-based mutagenesis of SigE verifies the importance of hydrophobic and electrostatic residues in type III chaperone function." Mol Microbiol **62**(4): 928-40.

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.

Konjufca, V., S. Y. Wanda, et al. (2006). "A recombinant attenuated Salmonella enterica serovar Typhimurium vaccine encoding Eimeria acervulina antigen offers protection against E. acervulina challenge." Infect Immun **74**(12): 6785-96.

Kubori, T. and J. E. Galan (2002). "Salmonella type III secretion-associated protein InvE controls translocation of effector proteins into host cells." Journal of Bacteriology **184**(17): 4699-4708.

Kubori, T., Y. Matsushima, et al. (1998). "Supramolecular structure of the Salmonella typhimurium type III protein secretion system." Science **280**(5363): 602-5.

Law, S. M., G. R. Bellomy, et al. (1993). "In vivo thermodynamic analysis of repression with and without looping in lac constructs. Estimates of free and local lac repressor concentrations and of physical properties of a region of supercoiled plasmid DNA in vivo." J Mol Biol **230**(1): 161-73.

Lazaris, A., S. Arcidiacono, et al. (2002). "Spider silk fibers spun from soluble recombinant silk produced in mammalian cells." Science **295**(5554): 472-6.

Lee, P. A., D. Tullman-Ercek, et al. (2006). "The bacterial twin-arginine translocation pathway." Annual Review of Microbiology **60**: 373-395.

Lee, S. H. and J. E. Galan (2004). "Salmonella type III secretion-associated chaperones confer secretion-pathway specificity." Mol Microbiol **51**(2): 483-95.

Lee, S. M., E. Pippel, et al. (2009). "Greatly increased toughness of infiltrated spider silk." Science **324**(5926): 488-92.

Lee, V. T. and O. Schneewind (2002). "Yop fusions to tightly folded protein domains and their effects on Yersinia enterocolitica type III secretion." J Bacteriol **184**(13): 3740-5.

Lewis, R. V. (2006). "Spider silk: ancient ideas for new biomaterials." Chem Rev **106**(9): 3762-74.

Lewis, R. V., M. Hinman, et al. (1996). "Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins." Protein Expr Purif **7**(4): 400-6.

Lucas, R. L. and C. A. Lee (2000). "Unravelling the mysteries of virulence gene regulation in Salmonella typhimurium." Mol Microbiol **36**(5): 1024-33.

Lucic, M. R., B. E. Forbes, et al. (1998). "Secretion in Escherichia coli and phage-display of recombinant insulin-like growth factor binding protein-2." J Biotechnol **61**(2): 95-108.

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

Lutz, R. and H. Bujard (1997). "Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements." Nucleic Acids Res **25**(6): 1203-10.

Majander, K., L. Anton, et al. (2005). "Extracellular secretion of polypeptides using a modified Escherichia coli flagellar secretion apparatus." Nat Biotechnol **23**(4): 475-81.

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

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

Mehta, R. A. and J. D. Kahn (1999). "Designed hyperstable Lac repressor.DNA loop topologies suggest alternative loop geometries." J Mol Biol **294**(1): 67-77.

Mossing, M. C. and M. T. Record, Jr. (1986). "Upstream operators enhance repression of the lac promoter." Science **233**(4766): 889-92.

Muller, J., S. Oehler, et al. (1996). "Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator." J Mol Biol **257**(1): 21-9.

Nielsen, H., J. Engelbrecht, et al. (1997). "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." Protein Engineering **10**(1): 1-6.

Pohlschroder, M., E. Hartmann, et al. (2005). "Diversity and evolution of protein translocation." Annu Rev Microbiol **59**: 91-111.

Prince, J. T., K. P. McGrath, et al. (1995). "Construction, cloning, and expression of synthetic genes encoding spider dragline silk." Biochemistry **34**(34): 10879-85.

Purnick, P. E. and R. Weiss (2009). "The second wave of synthetic biology: from modules to systems." Nat Rev Mol Cell Biol **10**(6): 410-22.

Randolph V. Lewis, M. H., Srinivas Kothakota, and Maurille J. Fournier (1996). "Expression and Purification of a Spider Silk Protein: A New Strategy for Producing Repetitive Proteins." Protein Expressing and Purification **7**: 6.

Rising, A., H. Nimmervoll, et al. (2005). "Spider silk proteins--mechanical property and gene sequence." Zoolog Sci **22**(3): 273-81.

Rodrigue, A., A. Chanal, et al. (1999). "Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway." Journal of Biological Chemistry **274**(19): 13223-13228.

Russmann, H., T. Kubori, et al. (2002). "Molecular and functional analysis of the type III secretion signal of the Salmonella enterica InvJ protein." Mol Microbiol **46**(3): 769-79.

Russmann, H., H. Shams, et al. (1998). "Delivery of epitopes by the Salmonella type III secretion system for vaccine development." Science **281**(5376): 565-8.

Sargent, F., E. G. Bogsch, et al. (1998). "Overlapping functions of components of a bacterial Sec-independent protein export pathway." Embo Journal **17**(13): 3640-3650.

Schlumberger, M. C., A. J. Muller, et al. (2005). "Real-time imaging of type III secretion: Salmonella SipA injection into host cells." Proc Natl Acad Sci U S A **102**(35): 12548-53.

Seidel, A., O. Liivak, et al. (2000). "Regenerated Spider Silk: Processing, Properties and Structure." Macromolecules **33**: 775-780.

Somerville, C. (2007). "Biofuels." Curr Biol **17**(4): R115-9.

Sorensen, H. P. and K. K. Mortensen (2005). "Advanced genetic strategies for recombinant protein expression in Escherichia coli." Journal of Biotechnology **115**(2): 113-128.

Stebbins, C. E. and J. E. Galan (2001). "Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion." Nature **414**(6859): 77-81.

Sutherland, T. D., J. H. Young, et al. (2009). "Insect Silk: One Name, Many Materials." Annu Rev Entomol.

Swanson, B. O., T. A. Blackledge, et al. (2006). "Spider dragline silk: correlated and mosaic evolution in high-performance biological materials." Evolution Int J Org Evolution **60**(12): 2539-51.

Tartera, C. and E. S. Metcalf (1993). "Osmolarity and Growth-Phase Overlap in Regulation of Salmonella-Typhi Adherence to and Invasion of Human Intestinal-Cells." Infection and Immunity **61**(7): 3084-3089.

Temme, K., H. Salis, et al. (2008). "Induction and Relaxation Dynamics of the Regulatory Network Controlling the Type III Secretion System encoded within Salmonella Pathogenicity Island 1." J. Mol Biol.

Temme, K., H. Salis, et al. (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**(1): 47-61.

Teule, F., A. R. Cooper, et al. (2009). "A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning." Nat Protoc **4**(3): 341-55.

Tong, L., Q. Lin, et al. (2000). "Extracellular expression, purification, and characterization of a winter flounder antifreeze polypeptide from Escherichia coli." Protein Expr Purif **18**(2): 175-81.

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

Vegh, B. M., P. Gal, et al. (2006). "Localization of the flagellum-specific secretion signal in Salmonella flagellin." Biochem Biophys Res Commun **345**(1): 93-8.

Vendrely, C. and T. Scheibel (2007). "Biotechnological production of spider-silk proteins enables new applications." Macromolecular Bioscience **7**(4): 401-409.

Villalobos, A., J. E. Ness, et al. (2006). "Gene Designer: a synthetic biology tool for constructing artificial DNA segments." BMC Bioinformatics **7**: 285.

Vollrath, F. and D. P. Knight (2001). "Liquid crystalline spinning of spider silk." Nature **410**(6828): 541-8.

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

Wang, Y., H. J. Kim, et al. (2006). "Stem cell-based tissue engineering with silk biomaterials." Biomaterials **27**(36): 6064-82.

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

Widmaier, D. M., D. Tullman-Ercek, et al. (2009). "Engineering the Salmonella type III secretion system to export spider silk monomers." Mol Syst Biol **5**: 309.

Wittrup, K. D. (2000). "The single cell as a microplate well." Nature Biotechnology **18**(10): 1039-1040.

Wood, M. W., C. Williams, et al. (2004). "Structural analysis of Salmonella enterica effector protein SopD." Biochim Biophys Acta **1698**(2): 219-26.

Xu, M. and R. V. Lewis (1990). "Structure of a protein superfiber: spider dragline silk." Proc Natl Acad Sci U S A **87**(18): 7120-4.

Yip, C. K., T. G. Kimbrough, et al. (2005). "Structural characterization of the molecular platform for type III secretion system assembly." Nature **435**(7042): 702-7.

Zhang, S., R. L. Santos, et al. (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**(7): 3843-55.

Zhang, Y. H. P., M. E. Himmel, et al. (2006). "Outlook for cellulase improvement: Screening and selection strategies." Biotechnology Advances **24**(5): 452-481.

Zhou, S. D., L. P. Yomano, et al. (1999). "Enhancement of expression and apparent secretion of Erwinia chrysanthemi endoglucanase (encoded by celZ) in Escherichia coli B." Applied and Environmental Microbiology **65**(6): 2439-2445.

Appendix A. Silk DNA and Amino Acid Sequences

A.1.1. *Latrodectus hesperus* Major Ampullate Spidroin 1 (LHF1), Genbank ascension EF595246 (Ayoub et al. 2007)

Synthetic DNA Sequence:

ATGACGTGGTCTACGCGCCTGGCACTGTCTTTCCTGTTTGTACTGTGCA
CTCAGTCCCTGTACGCACTCGCTCAGGCTAATACTCCGTGGTCTAGCAAGGCA
AACGCAGATGCATTCATTAATTCCTTTATCTCCGCAGCCTCTAACACGGGCTC
TTTTAGCCAAGACCAAATGGAGGACATGAGCCTGATCGGCAACACGCTGATG
GCTGCCATGGATAACATGGGCGGTCGCATCACGCCGTCTAAGCTGCAAGCGC
TGGACATGGCATTTGCATCCTCTGTGGCGGAGATCGCTGCCTCCGAGGGTGGT
GATCTGGGTGTGACGACTAATGCAATCGCAGACGCTCTGACTTCCGCGTTCTA
CCAAACTACTGGTGTGGTGAACAGCCGTTTTATTTCCGAGATCCGCTCTCTGA
TCGGTATGTTCGCACAAGCCTCCGCCAATGATGTCTACGCTTCCGCAGGTTCT
AGCGGTGGCGGCGGGTACGGGGCGTCTAGTGCTAGCGCAGCCTCAGCATCGG
CCGCAGCCCCGTCGGGAGTTGCTTATCAAGCTCCGGCACAGGCTCAAATTAG
CTTTACATTGCGTGGACAACAACCGGTATCTTACGGGCAAGGAGGCGCTGGG

CCAGGCGGTGCCGGGGCAGCTGCAGCCGCAGCGGCTGCTGCGGGCGGGGCA

GGACAGGGCGGGCAGGGAGGCTATGGGCAAGGCGGATACGGCCAGGGAGGT

GCAGGCCAAGGAGGGTCAGGAGCTGCCGCCGCGGCCGCTGCTGCAGCAGGT

GGAACGGGACAGGGTGGAGCAGGTCAGGGTGGCGCAGGTGCTGCCGCGGCT

GCCGCTGCCGCAGCTGGAGGAGCCGGACAAGGAGGTCAGGGTGGTTACGGG

CAGGGTGGTTATGGGCAGGGCGGAACAGGGCAGGGAGGTGCTGGCGCGGCC

GCCGCAGCCGCAGCAGCAGGTGGAGCGGGTCAGGGAGGGCAAGGAGGCTAC

GGTCAGGGCGGTTATGGCCAGGGAGGGTATGGCCAGGGTGGATCCGGAGCT

GCCGCAGCTGCAGCAGCGGCAGCCGGAGGCGCGGGCCAGGGCGGGCAGGGT

GGGTATGGACAGGGAGGTTACGGCCAGGGAGGCGCCGGACAAGGCGGTGCA

GGCGCTGCAGCCGCCGCTGCGGCCGCTGCAGGAGGAGCTGGACAAGGAGGC

TACGGCAGGGGCGGCGCGGGACAGGGTGGCGCTGCCGCGGCGGCAGCAGCC

GCTGCCGGTGCGGGCCAAGGCGGCTATGGCGGCCAGGGTGCAGGACAGGGT

GGCAGTGGGGCAGCTGCGGCGGCCGCGGCAGCTGGTGGGGCAGGCCAAGGA

GGGCAAGGCGGATACGGGCAAGGCGGGTATGGTCAAGGTGGCTCTGGTGCA

GCGGCCGCGGCTGCTGCTGCAGGAGGGGCAGGCCAGGGAGGGCAGGGTGGG

TATGGGCAGGGAGGATACGGCCAAGGAGGCGCGGGCCAGGGCGGAGCAGGG

GCAGCTGCTGCGGCAGCGGCTGCAGGCGGCGCCGGCCAGGGAGGTCAGGGA

GGATATGGACAGGGCGGGTACGGACAGGGTGGTGCGGGACAGGGCGGGGCG

GGCGCGGCGGCAGCGGCGGCGGCTGCCGGAGGGGCTGGGCAGGGTGGACAA

GGAGGTTATGGGCAGGGAGGCTATGGACAGGGAGGTGCTGGCCAGGGTGGC

GCAGGCGCTGCTGCCGCTGCGGCCGCAGCAGGCGGAGCCGGACAGGGTGGT

CAGGGTGGTTATGGACAAGGAGGCTACGGGCAGGGTGGAGCTGGCCAGGGC

GGGGCGGCTGCTGCGGCCGCGGCAGCAGCAGGAGGAGCGGGCCAAGGAGGA

TACGGGAGGGGAGGGGCGGGACAGGGTGGTGCGGCGGCCGCTGCTGGTGCA

GGACAAGGCGGGTACGGAGGCCAGGGTGCGGGTCAGGGCGGAGCAGGGGCT

GCGGCTGCAGCAGCTGCGGCGGGAGGTGCAGGCCAAGGCGGACAAGGAGGG

TATGGACGCGGAGGTTATGGGCAGGGCGGTGCTGGGCAAGGAGGGGCGGGC

GCGGCTGCCGCCGCCGCTGCCGCAGGAGGCGCAGGGCAGGGTGGACAGGGC

GGCTACGGTCAGGGCGGATATGGCCAAGGAGGGGCTGGACAAGGAGGAGCG

GCAGCAGCCGCAGCAGCAGCGGGAGGTGCGGGGCAAGGAGGGTACGGCAGG

GGCGGCGCAGGCCAGGGTGGAGCGGCGGCTGCCGCAGCCGCTGCGGCCGGA

GCGGGTCAGGGTGGCTATGGAGGTCAGGGTGCCGGTCAGGGAGGGGCTGGT

GCCGCGGCGGCAGCTGCAGCAGCTGGTGGCGCCGGGCAGGGCGGCCAAGGG

GATTACGGTAGAGGCGGTTACGGGCAAGGCGGAGCCGGCCAAGGCGGTGCC

GGAGCAGCGGCAGCTGCAGCCGCGGCCGGCGGAGCAGGCCAGGGCGGACAA

GGCGGTTATGGCCAAGGAGGATACGGACAAGGAGGCGCAGGACAGGGTGGT

GCGGCGGCGGCCGCAAGCGCAGCGGCAGCTGGCGGGGCCGGGCAGGGTGGG

TATGGACGCGGAGGCGCGGGTCAGGGTGGCGCGGCGGCTGCGGCCGGTGCG

GGGCAAGGAGGTTATGGAGGTCAGGGAGCGGGGCAAGGCGGAGCAGGAGCA

GCAGCCGCAGCTGCTGCGGCGGGAGGAGCCGGCCAAGGAGGGCAGGGCGGA

TATGGTCGAGGAGGGTACGGTCAAGGAGGCGCAGGACAGGGAGGGGCGGGT

GCTGCAGCCGCTGCAACCGCTGCGGGCGGGGCAGGCCAAGGCGGACAGGGA

GGTTACGGACAGGGCGGGTATGGCCAGGGTGGTGCTGGGCAAGGCGGGGCG

GCGGCAGCGGCCGCGGCTGCCGCAGGAGGTGCCGGCCAGGGCGGCTATGGA

CGGGGCGGTGCCGGACAGGGAGGAGCAGCGGCCGCGGCCGCTGCAGCAGCG

GGTGCCGGCCAAGGAGGGTATGGCGGACAGGGAGCCGGTCAGGGTGGCGCA

GGAGCCGCAGCCGCTGCCGCTGGTGGAGCTGGACAGGGTGGGCAGGGCGGA

TACGGCCGTGGCGGGTATGGTCAGGGAGGTGCTGGGCAAGGAGGAGCTGGA

GCAGCCGCCGCCGCAGCAGCTGCAGGCGGCGCCGGCCAGGGTGGTCAAGGA

GGTTACGGTCAAGGAGGATATGGCCAGGGTGGGGCAGGACAGGGCGGAGCT

GCCGCGGCTGCCGCCGCAGCTGCCGGTGGTGCGGGTCAAGGAGGTTATGGAC

GAGGTGGCGCGGGGCAGGGAGGAGCAGCGGCGGCCGCAGGCGCGGGCCAG

GGCGGGTACGGCGGGCAGGGCGCAGGCCAGGGCGGCGCGGGAGCTGCTGCT

GCCGCTTCAAGAGGTGCAGGTCAAGGTGGTCAAGGAGGGTATGGCCGCGGA

GGATATGGACAGGGTGGAGCTGGACAGGGAGGTGCGGGCGCGGCAGCCGCC

GCGGCTGCGGCTGGTGGAGCTGGCCAAGGCGGTCAAGGCGGCTATGGTCAGG

GCGGTTATGGGCAAGGAGGAGCCGGCCAGGGCGGGGCGGCGGCAGCGGCTG

CTGCGGCGGGCGGCGCAGGCCAAGGAGGATACGGTCGTGGCGGTGCCGGCC

AAGGAGGCGCAGCTGCCGCAGCGGGTGCGGGGCAGGGTGGCTACGGTGGGC

AGGGTGCCGGTCAGGGTGGTGCCGGAGCGGCCGCCGCTGCTGCTGCAGCTGG

TGGCGCGGGCCAGGGTGGACAGGGCGGATATGGCCGAGGAGGGTACGGACA

GGGCGGCGCAGGGCAAGGCGGGGCAGGAGCCGCCGCCGCTGCTGCGGCGGC

CGGAGGGGCTGGACAGGGTGGGCAGGGAGGCTATGGCCAGGGTGGTTATGG

TCAAGGAGGAGCCGGACAGGGTGGTGCTGCCGCTGCCGCCGCGGCTGCTGCC

GGTGGAGCCGGTCAGGGCGGCTACGGTCGAGGCGGTGCCGGCCAGGGCGGT

GCGGCCGCTGCCGCAGCCGCTGCAGCAGGAAGCGGACAAGGTGGATATGGC

GGTCAAGGCGCTGGGCAAGGAGGGGCCGGGGCGGCTGCAGCAGCAGCAGCT

GCCGGAGGCGCGGGTCAAGGCGGCCAAGGCGGATACGGGCGAGGCGGATAC

GGTCAGGGAGGTGCGGGGCAAGGCGGAGCAGGCGCGGCAGCCGCTGCGGCC

GCTGCGGGTGGGGCTGGTCAAGGAGGCCAGGGCGGATACGGCCAGGGTGGT

TACGGACAGGGAGGAGCAGGGCAAGGCGGAGCTGCAGCCGCGGCAGCCGCC

GCGGCCGCCGGAGGAGCGGGGCAGGGAGGGTACGGACGAGGAGGTGCTGGG

CAGGGCGGCGCTGCCGCCGCCGCTGGCGCCGGCCAAGGAGGATACGGTGGT

CAGGGTGCCGGACAGGGCGGAGCTGGAGCTGCCGCTGCAGCAGCCGCAGCA

GGTGGGGCGGGACAGGGAGGTCAAGGCGGCTATGGACGGGGCGGATACGGG

CAAGGAGGGGCTGGCCAGGGAGGGGCAGGTACTGCAGCTGCGGCAGCAGCG

GCTGGCGGAGCAGGCCAAGGAGGACAGGGTGGGTACGGTCAAGGCGGGTAT

GGGCAGGGTGGGGCTGGACAGGGCGGTGCAGCCGCGGCAGCCGCTGCAGCC

GCGGGCGGTGCGGGCCAAGGAGGTTACGGGCGAGGTGGTGCGGGTCAAGGC

GGGGCCGCTGCCGCAGCTGCCGCAGCGGCAGGAGCTGGGCAGGGAGGATAC

GGAGGGCAAGGGGCGGGTCAAGGCGGGGCTGGCGCGGCAGCCGCTGCTGCA

GCTGCTGGAGGAGCGGGGCAGGGTGGCCAAGGTGGTTATGGTCGCGGAGGTT

ACGGACAAGGAGGGGCCGGCCAGGGCGGCGCAGGAGCGGCGGCGGCTGCGG

CTGCAGCGGGTGGCGCGTCCCAGGGTGGACAAGGTGGGTACGGTCAAGGGG

ATTATGGGCAAGGAGGAGCAGGTCAGGGTGGGGCTGCTGCTGCGGCCGCGG

CGGCCGGTGGAGCAGGACAGGGAGGGTATGGACGTGGTGGCGCAGGCCAAG

GCGGAGCTGCTGCCGCTGCCGGAGCAGGACAAGGAGGTTACGGAGGGCAGG

GTGCTGGGCAGGGAGGAGCGGGTGCAGCGGCAGCTGCGGCGGCGGCTGGTG

GTGCCGGAAGGGGTGGGCAAGGCGGGTATGGACGTGGCGGATACGGGCAGG

GAGGGGCAGGCCAAGGTGGTGCGGGTGCTGCGGCCGCGGCGGCAGCGGCCG

GTGGTGCTGGCCAAGGTGGCCAGGGTGGCTACGGACAAGGTGGGTATGGTCA

AGGTGGAACTGGCCAAGGCGGTGCAGCTGCTGCTGCCGCGGCTGCTGCGGGC

GGAGCTGGACAAGGAGGCTATGGTCGGGGTGGAGCAGGGCAAGGAGGGGCT

GCAGCCGCTGCGGCTGCGGCAGCCGGGGCCGGTCAGGGTGGTTATGGAGGG

CAGGGCGCGGGTCAAGGCGGGGCTGGGGCGGCTGCTGCTGCAGCAGCGGCC

GGCGGCGCCGGGCAGGGTGGACAGGGTGGCTACGGAAGAGGAGGCTACGGT

CAAGGAGGCGCAGGCCAAGGCGGGGCAGGAGCTGCTGCTGCCGCGGCGGCC

GCGGGTGGTGCAGGGCAAGGAGGACAGGGAGGGTACGGCCAAGGAGGATAT

GGGCAAGGCGGTTACGGTCAGGGTGGAGCTGGACAAGGCGGAGCGGCTGCC

GCTGCTGCAGCTGCTGGCGGCGCGGGACAAGGTGGTTACGGGAGGGGAGGT

GCCGGACAGGGTGGCGCTGCAGCCGCCGCTGGTGCTGGCCAAGGAGGTTACG

GTGGCCAAGGTGCTGGGCAGGGTGGGGCTGGTGCGGCTGCTGCGGCGGCAGC

TGCGGGAGGAGCAGGGCAGGGCGGACAAGGTGGCTATGGCAGGGGCGGTTA

TGGACAGGGCGGCGCAGGACAAGGAGGAGCGGGCGCGGCGGCTGCAGCAGC

CGCTGCCGGAGGCGCCGGTCAGGGTGGCCAGGGCGGGTACGGCCAAGGCGG

ATACGGTCAAGGCGGGGCAGGACAGGGTGGCGCTGCGGCTGCGGCAGCAGC

AGCGGCAGGAGGGGCGGGCCAAGGCGGATATGGGCGGGGCGGAGCGGGAC

AAGGCGGAGCGGCAGCTGCAGCGGCCGCAGCGGCGGGTTCAGGCCAGGGAG

GTTACGGCGGCCAAGGTGCCGGACAGGGAGGAGCAGGCGCCGCGGCCGCCG

CAGCAGCTGCGGGAGGCGCCGGGCAGGGAGGGCAAGGAGGTTACGGGCGTG

GCGGCTACGGACAAGGTGGCGCGGGCCAGGGTGGGGCTGGCGCAGCGGCGG

CGGCTGCCGCCGCAGGCGGCGCTGGGCAAGGCGGGCAGGGTGGATACGGGC

AGGGCGGATATGGCCAGGGTGGGTACGGTCAAGGAGGTGCCGGACAGGGAG

GAGCAGCGGCAGCAGCTGCGGCGGCGGCAGCTGGAGGGGCAGGGCAGGGCG

GTTATGGTCGTGGAGGGGCCGGACAAGGAGGTGCAGCTGCAGCAGCGGGAG

CTGGGCAGGGTGGATACGGAGGTCAGGGCGCAGGACAGGGTGGTGCTGGCG

CTGCGGCTGCGGCAGCCGCGGCGGGAGGCGCGGGGCAGGGAGGCCAAGGCG

GGTATGGAAGAGGAGGCTATGGGCAAGGCGGTGCGGGACAAGGTGGGGCCG

GTGCTGCGGCGGCTGCAGCGGCTGCTGGAGGTGCTGGCCAAGGAGGGCAAG

GCGGATACGGTCAGGGTGGAAATGGGCAAGGCGGAGCGGGCCAGGGCGGCG

CGGCCGCTGCCGCGGCCGCAGCAGGCGGCGCGGGCCAAGGCGGGTACGGTC

GTGGCGGGGCCGGACAAGGCGGAGCTGCCGCCGCCGCTGCGGCCGCGGCTG

GGGCTGGGCAGGGTGGATATGGAGGGCAAGGCGCGGGACAGGGTGGTGCAG

GCGCGGCTGCCGCTGCTGCTGCGGCGGGCGGAGCAGGGCAGGGCGGCCAGG

GTGGATACGGGAGAGGTGGCTACGGTCAAGGAGGGGCCGGCCAAGGTGGTG

CGGGAGCAGCCGCTGCAGCTGCGGCCGCTGGTGGGGCTTCGCAGGGTGGCCA

GGGTGGTTACGGGCAAGGTGATTACGGGCAGGGAGGAGCGGGACAAGGTGG

GGCCGCCGCTGCGGCTGCGGCAGCAGGTGGCGCTGGACAAGGAGGGTATGG

ACGAGGAGGGGCAGGCCAAGGCGGGGCTGCCGCTGCTGCAGGGGCAGGTCA

AGGCGGGTACGGCGGACAAGGTGCAGGGCAAGGAGGTGCTGGAGCTGCGGC

TGCAGCAGCAGCTGCGGGTGGGGCAGGTAGAGGAGGCCAAGGAGGCTACGG

AAGAGGCGGATACGGTCAGGGCGGTGCGGGCCAGGGTGGTGCTGGTGCTGC

CGCAGCCGCCGCTGCTGCCGGTGGGGCTGGACAGGGTGGACAAGGTGGGTA

CGGACAAGGTGGCTACGGGCAGGGAGGCGCGGGTCAAGGAGGGGCGGCGGC

AGCCGCTGCAGCTGCCGCTGGTGGCGCCGGACAAGGAGGCTATGGCCGTGGT

GGGGCGGGGCAAGGCGGCGCGGCGGCCGCGGCGGGTGCGGGGCAGGGCGGC

TATGGAGGTCAAGGAGCCGGCCAAGGTGGGGCAGGCGCAGCAGCCGCTGCG

GCCGCTGCCGGAGGTGCTGGGCGAGGAGGGCAGGGTGGGTATGGTAGAGGC

GGGTACGGGCAGGGAGGCGCCGGACAGGGAGGGGCTGGTGCCGCTGCGGCT

GCGGCAGCCGCCGGAGGAGCGGGTCAGGGTGGCCAGGGCGGCTACGGTCAG

GGCGGGTATGGGCAGGGCGGGGCTGGACAAGGTGGCGCCGCGGCAGCTGCA

GCAGCTGCGGTGGGAGGGGCGGGTCAGGGTGGATACGGCCGAGGTGGGGCA

GGTCAAGGCGGTGCTGCGGCCGCGGCGGCCGCTGCAGCCGCTGGTTCTGGTC

AGGGAGGTTATGGAGGCCAAGGAGCAGGCCAAGGCGGCGCCGGGGCCGCTG

CCGCCGCCGCTGCTGCTGGCGGAGCAGGGCAGGGTGGACAGGGCGGTTACG

GTGGTGGAGGATACGGACAAGGAGGTGCCGGCCAAGGCGGAGCTGGAGCTG

CTGCGGCAGCAGCAGCCGCGGGAGGCGCTGGACAGGGTGGGCAGGGCGGCT

ACGGTCAGGGTGGATACGGTCAGGGCGGGGCGGGGCAGGGTGGAGCCGCAG

CGGCTGCAGCTGCGGCCGCGGGCGGAGCAGGGCAGGGAGGCTATGGTCGGG

GTGGAGCCGGACAAGGAGGCGCGGCGGCCGCTACCGGTGCTGGACAAGGTG

GCTATGGCGGTCAAGGAGCAGGACAAGGCGGAGCGGGCGCCGCTGCTGCAG

CGGCAGCTGCTGGTGGGGCCGGTCAGGGTGGCCAAGGAGGATACGGCAGGG

GTGGCTACGGCCAAGGCGGTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCG

CCGCCGCAGCCGCCGGAGGAGCTGGGCAAGGCGGACAAGGAGGGTATGGCC

AAGGTGGGTACGGGCAGGGGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGG

CTGCAGCCGCCGGTGGTGCTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCG

GTCAAGGAGGCGCCGCGGCAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGG

GCGGTTATGGAGGACAAGGCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGG

CAGCGGCAGCTGCGGGCGGCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTC

GAGGTGGCTACGGACAGGGAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAG

CGGCGGGAGGAGCCGGTCAGGGTGGGCAAGGCGGCTACGGTCAGGGTGGAT
ATGGCCAAGGTGGTGCAGGGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCG
CAGCCGCAGGCGGCGCGGGCCAAGGAGGATACGGCGGATATGGTCAGCAAG
GCGGAGCTGGAGCGGCTGCTGCAGCGGCTTCAGGACCGGGCCAGATCTACTA
CGGTCCGCAGTCCGTTGCCGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAG
CACCGGCCACTTCTGCACGCATTTCGTCGCACGCGTCCGCATTGCTTAGTAAT
GGGCCGACAAATCCTGCGTCCATCTCCAACGTAATCAGTAACGCAGTAAGCC
AAATTTCTAGCTCTAACCCAGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCC
CTACTTGAGCTCGTAACGGCGCTATTGACTATTATCGGCTCATCAAATATTGG
TAGTGTGAACTATGACAGTTCGGGCCAGTACGCACAAGTCGTTACTCAGTCA
GTTCAAAACGCCTTCGCC

Amino Acid Sequence:

MTWSTRLALSFLFVLCTQSLYALAQANTPWSSKANADAFINSFISAASNT
GSFSQDQMEDMSLIGNTLMAAMDNMGGRITPSKLQALDMAFASSVAEIAASEG
GDLGVTTNAIADALTSAFYQTTGVVNSRFISEIRSLIGMFAQASANDVYASAGSS
GGGGYGASSASAASASAAAPSGVAYQAPAQAQISFTLRGQQPVSYGQGGAGPG
GAGAAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGSGAAAAAAAAAGG
TGQGGAGQGGAGAAAAAAAAAGGAGQGGQGGYGQGGYGQGGTGQGGAGAA
AAAAAAGGAGQGGQGGYGQGGYGQGGYGQGGSGAAAAAAAAAGGAGQGGQ
GGYGQGGYGQGGAGQGGAGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAA
AAAAAGAGQGGYGGQGAGQGGSGAAAAAAAAGGAGQGGQGGYGQGGYGQG
GSGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAGAAAAAAAAGGAG
QGGQGGYGQGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQ

GGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAA

AAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAA

AGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQ

GGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAG

QGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGDYGRGGYGQGGAGQGGA

GAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAASAAAAGGAGQG

GYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGG

QGGYGRGGYGQGGAGQGGAGAAAAATAAGGAGQGGQGGYGQGGYGQGGAG

QGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQG

AGQGGAGAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAG

GAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQG

GAAAAAGAGQGGYGGQGAGQGGAGAAAAASRGAGQGGQGGYGRGGYGQGG

AGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAG

GAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGG

AGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGY

GQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGSGQGG

YGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAA

AAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAAGGAGQGGY

GRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAAGGAGQGGQG

GYGRGGYGQGGAGQGGAGTAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQG

GAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAG

QGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAG

GASQGGQGGYGQGDYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGA

AAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGG

AGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGTGQGGAAAAAAAAAG

GAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAA

AAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYG

QGGYGQGGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAG

AGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQG

GAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAG

QGGYGRGGAGQGGAAAAAAAAAGSGQGGYGGQGAGQGGAGAAAAAAAAGG

AGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGY

GQGGYGQGGAGQGGAAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAG

QGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGA

GAAAAAAAAGGAGQGGQGGYGQGGNGQGGAGQGGAAAAAAAAGGAGQGG

YGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAG

QGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGASQGGQGGYGQGDYGQG

GAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAG

QGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGG

AGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGG

AAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQG

GAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAA

VGGAGQGGYGRGGAGQGGAAAAAAAAAAGSGQGGYGGQGAGQGGAGAAAA

AAAAGGAGQGGQGGYGGGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGG

YGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAATGAGQ

GGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAG

AAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAGGAGQGGY

GRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQ

GGQGGYGRGGYGQGGAGQGGAGAAAAGGAGQGGQGGYGQGGYGQGGAGQ

GGAAAAAAAAAAGGAGQGGYGGYGQQGGAGAAAAAASGPGQIYYGPQSVAA

PAAAAASALAAPATSARISSHASALLSNGPTNPASISNVISNAVSQISSSNPGASAC

DVLVQALLELVTALLTIIGSSNIGSVNYDSSGQYAQVVTQSVQNAFA

A.1.2. 8B LHF1:

Synthetic DNA Sequence:

GGTGGGGCCGGTCAGGGTGGCCAAGGAGGATACGGCAGGGGTGGCTA

CGGCCAAGGCGGTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCGCCGCCGC

AGCCGCCGGAGGAGCTGGGCAAGGCGGACAAGGAGGGTATGGCCAAGGTGG

GTACGGGCAGGGGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGGCTGCAGC

CGCCGGTGGTGCTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCGGTCAAGG

AGGCGCCGCGGCAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGGGCGGTTA

TGGAGGACAAGGCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGGCAGCGGC

AGCTGCGGGCGGCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTCGAGGTGG

CTACGGACAGGGAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAGCGGCGGG

AGGAGCCGGTCAGGGTGGGCAAGGCGGCTACGGTCAGGGTGGATATGGCCA

AGGTGGTGCAGGGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCGCAGCCGC

AGGCGGCGCGGGCCAAGGAGGATACGGCGGATATGGTCAGCAAGGCGGAGC

TGGAGCGGCTGCTGCAGCGGCTTCAGGACCGGGCCAGATCTACTACGGTCCG

CAGTCCGTTGCCGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAGCACCGG

CCACTTCTGCACGCATTTCGTCGCACGCGTCCGCATTGCTTAGTAATGGGCCG
ACAAATCCTGCGTCCATCTCCAACGTAATCAGTAACGCAGTAAGCCAAATTT
CTAGCTCTAACCCAGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCCCTACTT
GAGCTCGTAACGGCGCTATTGACTATTATCGGCTCATCAAATATTGGTAGTGT
GAACTATGACAGTTCGGGCCAGTACGCACAAGTCGTTACTCAGTCAGTTCAA
AACGCCTTCGCC

Amino Acid Sequence:

GGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGG
YGQGGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAA
GAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQ
GGAGAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAAGGAGQG
GYGGYGQQGGAGAAAAAASGPGQIYYGPQSVAAPAAAAASALAAPATSARISS
HASALLSNGPTNPASISNVISNAVSQISSSNPGASACDVLVQALLELVTALLTIIGS
SNIGSVNYDSSGQYAQVVTQSVQNAFA

A.1.3. 12B LHF1:

Synthetic DNA Sequence:

GGCGGAGCAGGGCAGGGTGGACAGGGCGGTTACGGTGGTGGAGGATA
CGGACAAGGAGGTGCCGGCCAAGGCGGAGCTGGAGCTGCTGCGGCAGCAGC
AGCCGCGGGAGGCGCTGGACAGGGTGGGCAGGGCGGCTACGGTCAGGGTGG
ATACGGTCAGGGCGGGGCGGGGCAGGGTGGAGCCGCAGCGGCTGCAGCTGC
GGCCGCGGGCGGAGCAGGGCAGGGAGGCTATGGTCGGGGTGGAGCCGGACA

AGGAGGCGCGGCGGCCGCTACCGGTGCTGGACAAGGTGGCTATGGCGGTCA

AGGAGCAGGACAAGGCGGAGCGGGCGCCGCTGCTGCAGCGGCAGCTGCTGG

TGGGGCCGGTCAGGGTGGCCAAGGAGGATACGGCAGGGGTGGCTACGGCCA

AGGCGGTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCGCCGCCGCAGCCGC

CGGAGGAGCTGGGCAAGGCGGACAAGGAGGGTATGGCCAAGGTGGGTACGG

GCAGGGGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGGCTGCAGCCGCCGG

TGGTGCTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCGGTCAAGGAGGCGC

CGCGGCAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGGGCGGTTATGGAGG

ACAAGGCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGGCAGCGGCAGCTGC

GGGCGGCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTCGAGGTGGCTACGG

ACAGGGAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAGCGGCGGGAGGAGC

CGGTCAGGGTGGGCAAGGCGGCTACGGTCAGGGTGGATATGGCCAAGGTGG

TGCAGGGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCGCAGCCGCAGGCGG

CGCGGGCCAAGGAGGATACGGCGGATATGGTCAGCAAGGCGGAGCTGGAGC

GGCTGCTGCAGCGGCTTCAGGACCGGGCCAGATCTACTACGGTCCGCAGTCC

GTTGCCGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAGCACCGGCCACTTC

TGCACGCATTTCGTCGCACGCGTCCGCATTGCTTAGTAATGGGCCGACAAATC

CTGCGTCCATCTCCAACGTAATCAGTAACGCAGTAAGCCAAATTTCTAGCTCT

AACCCAGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCCCTACTTGAGCTCGT

AACGGCGCTATTGACTATTATCGGCTCATCAAATATTGGTAGTGTGAACTATG

ACAGTTCGGGCCAGTACGCACAAGTCGTTACTCAGTCAGTTCAAAACGCCTT

CGCC

Amino Acid Sequence:

GGAGQGGQGGYGGGGYGQGGAGQGGAGAAAAAAAAGGAGQGGGQGG
YGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAATGAGQ
GGYGGQGAGQGGAGAAAAAAAAGGAGQGGGQGGYGRGGYGQGGAGQGGAG
AAAAAAAAGGAGQGGGQGGYGQGGYGQGGAGQGGAAAAAAAAGGAGQGGY
GRGGAGQGGAAAAAAAAAGAGQGGYGGGQGAGQGGAGAAAAAAAAGGAGQ
GGQGGYGRGGYGQGGAGQGGAGAAAAGGAGQGGGQGGYGQGGYGQGGAGQ
GGAAAAAAAAAAGGAGQGGYGGYGQQGGAGAAAAAASGPGQIYYGPQSVAA
PAAAAASALAAPATSARISSHASALLSNGPTNPASISNVISNAVSQISSSNPGASAC
DVLVQALLELVTALLTIIGSSNIGSVNYDSSGQYAQVVTQSVQNAFA

A.1.4. 16B LHF1:

Synthetic DNA Sequence:

GGAGGTGCTGGGCGAGGAGGGCAGGGTGGGTATGGTAGAGGCGGGTA
CGGGCAGGGAGGCGCCGGACAGGGAGGGGCTGGTGCCGCTGCGGCTGCGGC
AGCCGCCGGAGGAGCGGGTCAGGGTGGCCAGGGCGGCTACGGTCAGGGCGG
GTATGGGCAGGGCGGGGCTGGACAAGGTGGCGCCGCGGCAGCTGCAGCAGC
TGCGGTGGGAGGGGCGGGTCAGGGTGGATACGGCCGAGGTGGGGCAGGTCA
AGGCGGTGCTGCGGCCGCGGCGGCCGCTGCAGCCGCTGGTTCTGGTCAGGGA
GGTTATGGAGGCCAAGGAGCAGGCCAAGGCGGCGCCGGGGCCGCTGCCGCC
GCCGCTGCTGCTGGCGGAGCAGGGCAGGGTGGACAGGGCGGTTACGGTGGT
GGAGGATACGGACAAGGAGGTGCCGGCCAAGGCGGAGCTGGAGCTGCTGCG
GCAGCAGCAGCCGCGGGAGGCGCTGGACAGGGTGGGCAGGGCGGCTACGGT
CAGGGTGGATACGGTCAGGGCGGGGCGGGGCAGGGTGGAGCCGCAGCGGCT

GCAGCTGCGGCCGCGGGCGGAGCAGGGCAGGGAGGCTATGGTCGGGGTGGA

GCCGGACAAGGAGGCGCGGCGGCCGCTACCGGTGCTGGACAAGGTGGCTAT

GGCGGTCAAGGAGCAGGACAAGGCGGAGCGGGCGCCGCTGCTGCAGCGGCA

GCTGCTGGTGGGGCCGGTCAGGGTGGCCAAGGAGGATACGGCAGGGGTGGC

TACGGCCAAGGCGGTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCGCCGCC

GCAGCCGCCGGAGGAGCTGGGCAAGGCGGACAAGGAGGGTATGGCCAAGGT

GGGTACGGGCAGGGGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGGCTGCA

GCCGCCGGTGGTGCTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCGGTCAA

GGAGGCGCCGCGGCAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGGGCGGT

TATGGAGGACAAGGCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGGCAGCG

GCAGCTGCGGGCGGCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTCGAGGT

GGCTACGGACAGGGAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAGCGGCG

GGAGGAGCCGGTCAGGGTGGGCAAGGCGGCTACGGTCAGGGTGGATATGGC

CAAGGTGGTGCAGGGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCGCAGCC

GCAGGCGGCGCGGGCCAAGGAGGATACGGCGGATATGGTCAGCAAGGCGGA

GCTGGAGCGGCTGCTGCAGCGGCTTCAGGACCGGGCCAGATCTACTACGGTC

CGCAGTCCGTTGCCGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAGCACC

GGCCACTTCTGCACGCATTTCGTCGCACGCGTCCGCATTGCTTAGTAATGGGC

CGACAAATCCTGCGTCCATCTCCAACGTAATCAGTAACGCAGTAAGCCAAAT

TTCTAGCTCTAACCCAGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCCCTAC

TTGAGCTCGTAACGGCGCTATTGACTATTATCGGCTCATCAAATATTGGTAGT

GTGAACTATGACAGTTCGGGCCAGTACGCACAAGTCGTTACTCAGTCAGTTC

AAAACGCCTTCGCC

Amino Acid Sequence:

GGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGG
YGQGGYGQGGAGQGGAAAAAAAAVGGAGQGGYGRGGAGQGGAAAAAAAA
AAGSGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGGGGYGQGGAG
QGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGG
AGQGGYGRGGAGQGGAAAATGAGQGGYGGQGAGQGGAGAAAAAAAAGGAG
QGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQ
GGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYG
GQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAA
GGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAAGGAGQGGYGGYGQQ
GGAGAAAAAASGPGQIYYGPQSVAAPAAAAASALAAPATSARISSHASALLSNG
PTNPASISNVISNAVSQISSSNPGASACDVLVQALLELVTALLTIIGSSNIGSVNYDS
SGQYAQVVTQSVQNAFA

A.1.5. 20B LHF1:

Synthetic DNA Sequence:

GGTGGGGCAGGTAGAGGAGGCCAAGGAGGCTACGGAAGAGGCGGAT
ACGGTCAGGGCGGTGCGGGCCAGGGTGGTGCTGGTGCTGCCGCAGCCGCCGC
TGCTGCCGGTGGGGCTGGACAGGGTGGACAAGGTGGGTACGGACAAGGTGG
CTACGGGCAGGGAGGCGCGGGTCAAGGAGGGGCGGCGGCAGCCGCTGCAGC
TGCCGCTGGTGGCGCCGGACAAGGAGGCTATGGCCGTGGTGGGGCGGGGCA
AGGCGGCGCGGCGGCCGCGGCGGGTGCGGGGCAGGGCGGCTATGGAGGTCA
AGGAGCCGGCCAAGGTGGGGCAGGCGCAGCAGCCGCTGCGGCCGCTGCCGG
AGGTGCTGGGCGAGGAGGGCAGGGTGGGTATGGTAGAGGCGGGTACGGGCA

GGGAGGCGCCGGACAGGGAGGGGCTGGTGCCGCTGCGGCTGCGGCAGCCGC

CGGAGGAGCGGGTCAGGGTGGCCAGGGCGGCTACGGTCAGGGCGGGTATGG

GCAGGGCGGGGCTGGACAAGGTGGCGCCGCGGCAGCTGCAGCAGCTGCGGT

GGGAGGGGCGGGTCAGGGTGGATACGGCCGAGGTGGGGCAGGTCAAGGCGG

TGCTGCGGCCGCGGCGGCCGCTGCAGCCGCTGGTTCTGGTCAGGGAGGTTAT

GGAGGCCAAGGAGCAGGCCAAGGCGGCGCCGGGGCCGCTGCCGCCGCCGCT

GCTGCTGGCGGAGCAGGGCAGGGTGGACAGGGCGGTTACGGTGGTGGAGGA

TACGGACAAGGAGGTGCCGGCCAAGGCGGAGCTGGAGCTGCTGCGGCAGCA

GCAGCCGCGGGAGGCGCTGGACAGGGTGGGCAGGGCGGCTACGGTCAGGGT

GGATACGGTCAGGGCGGGGCGGGGCAGGGTGGAGCCGCAGCGGCTGCAGCT

GCGGCCGCGGGCGGAGCAGGGCAGGGAGGCTATGGTCGGGGTGGAGCCGGA

CAAGGAGGCGCGGCGGCCGCTACCGGTGCTGGACAAGGTGGCTATGGCGGT

CAAGGAGCAGGACAAGGCGGAGCGGGCGCCGCTGCTGCAGCGGCAGCTGCT

GGTGGGGCCGGTCAGGGTGGCCAAGGAGGATACGGCAGGGGTGGCTACGGC

CAAGGCGGTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCGCCGCCGCAGCC

GCCGGAGGAGCTGGGCAAGGCGGACAAGGAGGGTATGGCCAAGGTGGGTAC

GGGCAGGGGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGGCTGCAGCCGCC

GGTGGTGCTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCGGTCAAGGAGGC

GCCGCGGCAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGGGCGGTTATGGA

GGACAAGGCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGGCAGCGGCAGCT

GCGGGCGGCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTCGAGGTGGCTAC

GGACAGGGAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAGCGGCGGGAGGA

GCCGGTCAGGGTGGGCAAGGCGGCTACGGTCAGGGTGGATATGGCCAAGGT

GGTGCAGGGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCGCAGCCGCAGGC
GGCGCGGGCCAAGGAGGATACGGCGGATATGGTCAGCAAGGCGGAGCTGGA
GCGGCTGCTGCAGCGGCTTCAGGACCGGGCCAGATCTACTACGGTCCGCAGT
CCGTTGCCGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAGCACCGGCCACT
TCTGCACGCATTTCGTCGCACGCGTCCGCATTGCTTAGTAATGGGCCGACAAA
TCCTGCGTCCATCTCCAACGTAATCAGTAACGCAGTAAGCCAAATTTCTAGCT
CTAACCCAGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCCCTACTTGAGCTC
GTAACGGCGCTATTGACTATTATCGGCTCATCAAATATTGGTAGTGTGAACTA
TGACAGTTCGGGCCAGTACGCACAAGTCGTTACTCAGTCAGTTCAAAACGCC
TTCGCC

Amino Acid Sequence:

GGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGG
YGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAG
QGGYGGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAG
AAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAVGGAGQGG
YGRGGAGQGGAAAAAAAAAAGSGQGGYGGQGAGQGGAGAAAAAAAAGGAG
QGGQGGYGGGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQ
GGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAATGAGQGGYGGQGA
GQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAA
GGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQG
GAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGR
GGYGQGGAGQGGAGAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAA
AAAAGGAGQGGYGGYGQQGGAGAAAAAASGPGQIYYGPQSVAAPAAAAASA

LAAPATSARISSHASALLSNGPTNPASISNVISNAVSQISSSNPGASACDVLVQALL
ELVTALLTIIGSSNIGSVNYDSSGQYAQVVTQSVQNAFA

A.1.6. 24B LHF1:

Synthetic DNA Sequence:

GCGGGCGGAGCAGGGCAGGGCGGCCAGGGTGGATACGGGAGAGGTG
GCTACGGTCAAGGAGGGGCCGGCCAAGGTGGTGCGGGAGCAGCCGCTGCAG
CTGCGGCCGCTGGTGGGGCTTCGCAGGGTGGCCAGGGTGGTTACGGGCAAGG
TGATTACGGGCAGGGAGGAGCGGGACAAGGTGGGGCCGCCGCTGCGGCTGC
GGCAGCAGGTGGCGCTGGACAAGGAGGGTATGGACGAGGAGGGGCAGGCCA
AGGCGGGGCTGCCGCTGCTGCAGGGGCAGGTCAAGGCGGGTACGGCGGACA
AGGTGCAGGGCAAGGAGGTGCTGGAGCTGCGGCTGCAGCAGCAGCTGCGGG
TGGGGCAGGTAGAGGAGGCCAAGGAGGCTACGGAAGAGGCGGATACGGTCA
GGGCGGTGCGGGCCAGGGTGGTGCTGGTGCTGCCGCAGCCGCCGCTGCTGCC
GGTGGGGCTGGACAGGGTGGACAAGGTGGGTACGGACAAGGTGGCTACGGG
CAGGGAGGCGCGGGTCAAGGAGGGGCGGCGGCAGCCGCTGCAGCTGCCGCT
GGTGGCGCCGGACAAGGAGGCTATGGCCGTGGTGGGGCGGGGCAAGGCGGC
GCGGCGGCCGCGGCGGGTGCGGGGCAGGGCGGCTATGGAGGTCAAGGAGCC
GGCCAAGGTGGGGCAGGCGCAGCAGCCGCTGCGGCCGCTGCCGGAGGTGCT
GGGCGAGGAGGGCAGGGTGGGTATGGTAGAGGCGGGTACGGGCAGGGAGGC
GCCGGACAGGGAGGGGCTGGTGCCGCTGCGGCTGCGGCAGCCGCCGGAGGA
GCGGGTCAGGGTGGCCAGGGCGGCTACGGTCAGGGCGGGTATGGGCAGGGC
GGGGCTGGACAAGGTGGCGCCGCGGCAGCTGCAGCAGCTGCGGTGGGAGGG
GCGGGTCAGGGTGGATACGGCCGAGGTGGGGCAGGTCAAGGCGGTGCTGCG

GCCGCGGCGGCCGCTGCAGCCGCTGGTTCTGGTCAGGGAGGTTATGGAGGCC

AAGGAGCAGGCCAAGGCGGCGCCGGGGCCGCTGCCGCCGCCGCTGCTGCTG

GCGGAGCAGGGCAGGGTGGACAGGGCGGTTACGGTGGTGGAGGATACGGAC

AAGGAGGTGCCGGCCAAGGCGGAGCTGGAGCTGCTGCGGCAGCAGCAGCCG

CGGGAGGCGCTGGACAGGGTGGGCAGGGCGGCTACGGTCAGGGTGGATACG

GTCAGGGCGGGGCGGGGCAGGGTGGAGCCGCAGCGGCTGCAGCTGCGGCCG

CGGGCGGAGCAGGGCAGGGAGGCTATGGTCGGGGTGGAGCCGGACAAGGAG

GCGCGGCGGCCGCTACCGGTGCTGGACAAGGTGGCTATGGCGGTCAAGGAG

CAGGACAAGGCGGAGCGGGCGCCGCTGCTGCAGCGGCAGCTGCTGGTGGGG

CCGGTCAGGGTGGCCAAGGAGGATACGGCAGGGGTGGCTACGGCCAAGGCG

GTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCGCCGCCGCAGCCGCCGGAG

GAGCTGGGCAAGGCGGACAAGGAGGGTATGGCCAAGGTGGGTACGGGCAGG

GGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGGCTGCAGCCGCCGGTGGTG

CTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCGGTCAAGGAGGCGCCGCGG

CAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGGGCGGTTATGGAGGACAAG

GCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGGCAGCGGCAGCTGCGGGCG

GCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTCGAGGTGGCTACGGACAGG

GAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAGCGGCGGGAGGAGCCGGTC

AGGGTGGGCAAGGCGGCTACGGTCAGGGTGGATATGGCCAAGGTGGTGCAG

GGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCGCAGCCGCAGGCGGCGCGG

GCCAAGGAGGATACGGCGGATATGGTCAGCAAGGCGGAGCTGGAGCGGCTG

CTGCAGCGGCTTCAGGACCGGGCCAGATCTACTACGGTCCGCAGTCCGTTGC

CGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAGCACCGGCCACTTCTGCAC

GCATTTCGTCGCACGCGTCCGCATTGCTTAGTAATGGGCCGACAAATCCTGCG
TCCATCTCCAACGTAATCAGTAACGCAGTAAGCCAAATTTCTAGCTCTAACCC
AGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCCCTACTTGAGCTCGTAACGG
CGCTATTGACTATTATCGGCTCATCAAATATTGGTAGTGTGAACTATGACAGT
TCGGGCCAGTACGCACAAGTCGTTACTCAGTCAGTTCAAAACGCCTTCGCC

Amino Acid Sequence:

AGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGASQGGQGG
YGQGDYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQ
GGYGGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGA
AAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGY
GRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGRGGQGG
YGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQG
GAAAAAAAAVGGAGQGGYGRGGAGQGGAAAAAAAAAAGSGQGGYGGQGAG
QGGAGAAAAAAAAGGAGQGGQGGYGGGGYGQGGAGQGGAGAAAAAAAAG
GAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQG
GAAAATGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQ
GGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAA
AGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAA
AAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAGGAGQGGQGGYGQG
GYGQGGAGQGGAAAAAAAAAAGGAGQGGYGGYGQQGGAGAAAAAASGPGQ
IYYGPQSVAAPAAAAASALAAPATSARISSHASALLSNGPTNPASISNVISNAVSQI
SSSNPGASACDVLVQALLELVTALLTIIGSSNIGSVNYDSSGQYAQVVTQSVQNA
FA

A.1.7. 28B LHF1:

Synthetic DNA Sequence:

GGAGGCGCGGGGCAGGGAGGCCAAGGCGGGTATGGAAGAGGAGGCT
ATGGGCAAGGCGGTGCGGGACAAGGTGGGGCCGGTGCTGCGGCGGCTGCAG
CGGCTGCTGGAGGTGCTGGCCAAGGAGGGCAAGGCGGATACGGTCAGGGTG
GAAATGGGCAAGGCGGAGCGGGCCAGGGCGGCGCGGCCGCTGCCGCGGCCG
CAGCAGGCGGCGCGGGCCAAGGCGGGTACGGTCGTGGCGGGGCCGGACAAG
GCGGAGCTGCCGCCGCCGCTGCGGCCGCGGCTGGGGCTGGGCAGGGTGGAT
ATGGAGGGCAAGGCGCGGGACAGGGTGGTGCAGGCGCGGCTGCCGCTGCTG
CTGCGGCGGGCGGAGCAGGGCAGGGCGGCCAGGGTGGATACGGGAGAGGTG
GCTACGGTCAAGGAGGGGCCGGCCAAGGTGGTGCGGGAGCAGCCGCTGCAG
CTGCGGCCGCTGGTGGGGCTTCGCAGGGTGGCCAGGGTGGTTACGGGCAAGG
TGATTACGGGCAGGGAGGAGCGGGACAAGGTGGGGCCGCCGCTGCGGCTGC
GGCAGCAGGTGGCGCTGGACAAGGAGGGTATGGACGAGGAGGGGCAGGCCA
AGGCGGGGCTGCCGCTGCTGCAGGGGCAGGTCAAGGCGGGTACGGCGGACA
AGGTGCAGGGCAAGGAGGTGCTGGAGCTGCGGCTGCAGCAGCAGCTGCGGG
TGGGGCAGGTAGAGGAGGCCAAGGAGGCTACGGAAGAGGCGGATACGGTCA
GGGCGGTGCGGGCCAGGGTGGTGCTGGTGCTGCCGCAGCCGCCGCTGCTGCC
GGTGGGGCTGGACAGGGTGGACAAGGTGGGTACGGACAAGGTGGCTACGGG
CAGGGAGGCGCGGGTCAAGGAGGGGCGGCGGCAGCCGCTGCAGCTGCCGCT
GGTGGCGCCGGACAAGGAGGCTATGGCCGTGGTGGGGCGGGGCAAGGCGGC
GCGGCGGCCGCGGCGGGTGCGGGGCAGGGCGGCTATGGAGGTCAAGGAGCC
GGCCAAGGTGGGGCAGGCGCAGCAGCCGCTGCGGCCGCTGCCGGAGGTGCT

GGGCGAGGAGGGCAGGGTGGGTATGGTAGAGGCGGGTACGGGCAGGGAGGC

GCCGGACAGGGAGGGGCTGGTGCCGCTGCGGCTGCGGCAGCCGCCGGAGGA

GCGGGTCAGGGTGGCCAGGGCGGCTACGGTCAGGGCGGGTATGGGCAGGGC

GGGGCTGGACAAGGTGGCGCCGCGGCAGCTGCAGCAGCTGCGGTGGGAGGG

GCGGGTCAGGGTGGATACGGCCGAGGTGGGGCAGGTCAAGGCGGTGCTGCG

GCCGCGGCGGCCGCTGCAGCCGCTGGTTCTGGTCAGGGAGGTTATGGAGGCC

AAGGAGCAGGCCAAGGCGGCGCCGGGGCCGCTGCCGCCGCCGCTGCTGCTG

GCGGAGCAGGGCAGGGTGGACAGGGCGGTTACGGTGGTGGAGGATACGGAC

AAGGAGGTGCCGGCCAAGGCGGAGCTGGAGCTGCTGCGGCAGCAGCAGCCG

CGGGAGGCGCTGGACAGGGTGGGCAGGGCGGCTACGGTCAGGGTGGATACG

GTCAGGGCGGGGCGGGGCAGGGTGGAGCCGCAGCGGCTGCAGCTGCGGCCG

CGGGCGGAGCAGGGCAGGGAGGCTATGGTCGGGGTGGAGCCGGACAAGGAG

GCGCGGCGGCCGCTACCGGTGCTGGACAAGGTGGCTATGGCGGTCAAGGAG

CAGGACAAGGCGGAGCGGGCGCCGCTGCTGCAGCGGCAGCTGCTGGTGGGG

CCGGTCAGGGTGGCCAAGGAGGATACGGCAGGGGTGGCTACGGCCAAGGCG

GTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCGCCGCCGCAGCCGCCGGAG

GAGCTGGGCAAGGCGGACAAGGAGGGTATGGCCAAGGTGGGTACGGGCAGG

GGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGGCTGCAGCCGCCGGTGGTG

CTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCGGTCAAGGAGGCGCCGCGG

CAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGGGCGGTTATGGAGGACAAG

GCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGGCAGCGGCAGCTGCGGGCG

GCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTCGAGGTGGCTACGGACAGG

GAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAGCGGCGGGAGGAGCCGGTC

AGGGTGGGCAAGGCGGCTACGGTCAGGGTGGATATGGCCAAGGTGGTGCAG

GGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCGCAGCCGCAGGCGGCGCGG

GCCAAGGAGGATACGGCGGATATGGTCAGCAAGGCGGAGCTGGAGCGGCTG

CTGCAGCGGCTTCAGGACCGGGCCAGATCTACTACGGTCCGCAGTCCGTTGC

CGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAGCACCGGCCACTTCTGCAC

GCATTTCGTCGCACGCGTCCGCATTGCTTAGTAATGGGCCGACAAATCCTGCG

TCCATCTCCAACGTAATCAGTAACGCAGTAAGCCAAATTTCTAGCTCTAACCC

AGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCCCTACTTGAGCTCGTAACGG

CGCTATTGACTATTATCGGCTCATCAAATATTGGTAGTGTGAACTATGACAGT

TCGGGCCAGTACGCACAAGTCGTTACTCAGTCAGTTCAAAACGCCTTCGCC

Amino Acid Sequence:

GGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGG

YGQGGNGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAA

GAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQ

GGAGAAAAAAAAGGASQGGQGGYGQGDYGQGGAGQGGAAAAAAAAGGAGQ

GGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGRGG

QGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGA

GQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQ

GGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGA

GQGGQGGYGQGGYGQGGAGQGGAAAAAAAAVGGAGQGGYGRGGAGQGGA

AAAAAAAAAGSGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGGGG

YGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAA

AAAAAGGAGQGGYGRGGAGQGGAAAATGAGQGGYGGQGAGQGGAGAAAAA

AAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGY

GQGGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAG

AGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQG

GAGAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAAGGAGQGG

YGGYGQQGGAGAAAAAASGPGQIYYGPQSVAAPAAAAASALAAPATSARISSH

ASALLSNGPTNPASISNVISNAVSQISSSNPGASACDVLVQALLELVTALLTIIGSS

NIGSVNYDSSGQYAQVVTQSVQNAFA

A.1.8. 32B LHF1:

Synthetic DNA Sequence:

GGAGGCGCCGGGCAGGGAGGGCAAGGAGGTTACGGGCGTGGCGGCTA

CGGACAAGGTGGCGCGGGCCAGGGTGGGGCTGGCGCAGCGGCGGCGGCTGC

CGCCGCAGGCGGCGCTGGGCAAGGCGGGCAGGGTGGATACGGGCAGGGCGG

ATATGGCCAGGGTGGGTACGGTCAAGGAGGTGCCGGACAGGGAGGAGCAGC

GGCAGCAGCTGCGGCGGCGGCAGCTGGAGGGGCAGGGCAGGGCGGTTATGG

TCGTGGAGGGGCCGGACAAGGAGGTGCAGCTGCAGCAGCGGGAGCTGGGCA

GGGTGGATACGGAGGTCAGGGCGCAGGACAGGGTGGTGCTGGCGCTGCGGC

TGCGGCAGCCGCGGCGGGAGGCGCGGGGCAGGGAGGCCAAGGCGGGTATGG

AAGAGGAGGCTATGGGCAAGGCGGTGCGGGACAAGGTGGGGCCGGTGCTGC

GGCGGCTGCAGCGGCTGCTGGAGGTGCTGGCCAAGGAGGGCAAGGCGGATA

CGGTCAGGGTGGAAATGGGCAAGGCGGAGCGGGCCAGGGCGGCGCGGCCGC

TGCCGCGGCCGCAGCAGGCGGCGCGGGCCAAGGCGGGTACGGTCGTGGCGG

GGCCGGACAAGGCGGAGCTGCCGCCGCCGCTGCGGCCGCGGCTGGGGCTGG

GCAGGGTGGATATGGAGGGCAAGGCGCGGGACAGGGTGGTGCAGGCGCGGC

TGCCGCTGCTGCTGCGGCGGGCGGAGCAGGGCAGGGCGGCCAGGGTGGATA

CGGGAGAGGTGGCTACGGTCAAGGAGGGGCCGGCCAAGGTGGTGCGGGAGC

AGCCGCTGCAGCTGCGGCCGCTGGTGGGGCTTCGCAGGGTGGCCAGGGTGGT

TACGGGCAAGGTGATTACGGGCAGGGAGGAGCGGGACAAGGTGGGGCCGCC

GCTGCGGCTGCGGCAGCAGGTGGCGCTGGACAAGGAGGGTATGGACGAGGA

GGGGCAGGCCAAGGCGGGGCTGCCGCTGCTGCAGGGGCAGGTCAAGGCGGG

TACGGCGGACAAGGTGCAGGGCAAGGAGGTGCTGGAGCTGCGGCTGCAGCA

GCAGCTGCGGGTGGGGCAGGTAGAGGAGGCCAAGGAGGCTACGGAAGAGGC

GGATACGGTCAGGGCGGTGCGGGCCAGGGTGGTGCTGGTGCTGCCGCAGCCG

CCGCTGCTGCCGGTGGGGCTGGACAGGGTGGACAAGGTGGGTACGGACAAG

GTGGCTACGGGCAGGGAGGCGCGGGTCAAGGAGGGGCGGCGGCAGCCGCTG

CAGCTGCCGCTGGTGGCGCCGGACAAGGAGGCTATGGCCGTGGTGGGGCGG

GGCAAGGCGGCGCGGCGGCCGCGGCGGGTGCGGGGCAGGGCGGCTATGGAG

GTCAAGGAGCCGGCCAAGGTGGGGCAGGCGCAGCAGCCGCTGCGGCCGCTG

CCGGAGGTGCTGGGCGAGGAGGGCAGGGTGGGTATGGTAGAGGCGGGTACG

GGCAGGGAGGCGCCGGACAGGGAGGGGCTGGTGCCGCTGCGGCTGCGGCAG

CCGCCGGAGGAGCGGGTCAGGGTGGCCAGGGCGGCTACGGTCAGGGCGGGT

ATGGGCAGGGCGGGGCTGGACAAGGTGGCGCCGCGGCAGCTGCAGCAGCTG

CGGTGGGAGGGGCGGGTCAGGGTGGATACGGCCGAGGTGGGGCAGGTCAAG

GCGGTGCTGCGGCCGCGGCGGCCGCTGCAGCCGCTGGTTCTGGTCAGGGAGG

TTATGGAGGCCAAGGAGCAGGCCAAGGCGGCGCCGGGGCCGCTGCCGCCGC

CGCTGCTGCTGGCGGAGCAGGGCAGGGTGGACAGGGCGGTTACGGTGGTGG

AGGATACGGACAAGGAGGTGCCGGCCAAGGCGGAGCTGGAGCTGCTGCGGC

AGCAGCAGCCGCGGGAGGCGCTGGACAGGGTGGGCAGGGCGGCTACGGTCA

GGGTGGATACGGTCAGGGCGGGGCGGGGCAGGGTGGAGCCGCAGCGGCTGC

AGCTGCGGCCGCGGGCGGAGCAGGGCAGGGAGGCTATGGTCGGGGTGGAGC

CGGACAAGGAGGCGCGGCGGCCGCTACCGGTGCTGGACAAGGTGGCTATGG

CGGTCAAGGAGCAGGACAAGGCGGAGCGGGCGCCGCTGCTGCAGCGGCAGC

TGCTGGTGGGGCCGGTCAGGGTGGCCAAGGAGGATACGGCAGGGGTGGCTA

CGGCCAAGGCGGTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCGCCGCCGC

AGCCGCCGGAGGAGCTGGGCAAGGCGGACAAGGAGGGTATGGCCAAGGTGG

GTACGGGCAGGGGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGGCTGCAGC

CGCCGGTGGTGCTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCGGTCAAGG

AGGCGCCGCGGCAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGGGCGGTTA

TGGAGGACAAGGCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGGCAGCGGC

AGCTGCGGGCGGCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTCGAGGTGG

CTACGGACAGGGAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAGCGGCGGG

AGGAGCCGGTCAGGGTGGGCAAGGCGGCTACGGTCAGGGTGGATATGGCCA

AGGTGGTGCAGGGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCGCAGCCGC

AGGCGGCGCGGGCCAAGGAGGATACGGCGGATATGGTCAGCAAGGCGGAGC

TGGAGCGGCTGCTGCAGCGGCTTCAGGACCGGGCCAGATCTACTACGGTCCG

CAGTCCGTTGCCGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAGCACCGG

CCACTTCTGCACGCATTTCGTCGCACGCGTCCGCATTGCTTAGTAATGGGCCG

ACAAATCCTGCGTCCATCTCCAACGTAATCAGTAACGCAGTAAGCCAAATTT

CTAGCTCTAACCCAGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCCCTACTT

GAGCTCGTAACGGCGCTATTGACTATTATCGGCTCATCAAATATTGGTAGTGT

GAACTATGACAGTTCGGGCCAGTACGCACAAGTCGTTACTCAGTCAGTTCAA
AACGCCTTCGCC

Amino Acid Sequence:

GGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGG
YGQGGYGQGGYGQGGAGQGGAAAAAAAAAAGGAGQGGYGRGGAGQGGAA
AAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGG
AGQGGAGAAAAAAAAGGAGQGGQGGYGQGGNGQGGAGQGGAAAAAAAAG
GAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAA
AAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGASQGGQGGYGQ
GDYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGY
GGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGAAAA
AAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRG
GAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGR
GGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAA
AAAAAAVGGAGQGGYGRGGAGQGGAAAAAAAAAAGSGQGGYGGQGAGQGG
AGAAAAAAAAGGAGQGGQGGYGGGGYGQGGAGQGGAGAAAAAAAAGGAG
QGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAA
AATGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGA
GQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAGG
AGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAA
AGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAGGAGQGGQGGYGQGGYG
QGGAGQGGAAAAAAAAAAGGAGQGGYGGYGQQGGAGAAAAAASGPGQIYY

GPQSVAAPAAAAASALAAPATSARISSHASALLSNGPTNPASISNVISNAVSQISSS

NPGASACDVLVQALLELVTALLTIIGSSNIGSVNYDSSGQYAQVVTQSVQNAFA

A.1.9. 36B LHF1:

Synthetic DNA Sequence:

GGAGGAGCAGGGCAGGGCGGACAAGGTGGCTATGGCAGGGGCGGTTA

TGGACAGGGCGGCGCAGGACAAGGAGGAGCGGGCGCGGCGGCTGCAGCAGC

CGCTGCCGGAGGCGCCGGTCAGGGTGGCCAGGGCGGGTACGGCCAAGGCGG

ATACGGTCAAGGCGGGGCAGGACAGGGTGGCGCTGCGGCTGCGGCAGCAGC

AGCGGCAGGAGGGGCGGGCCAAGGCGGATATGGGCGGGGCGGAGCGGGAC

AAGGCGGAGCGGCAGCTGCAGCGGCCGCAGCGGCGGGTTCAGGCCAGGGAG

GTTACGGCGGCCAAGGTGCCGGACAGGGAGGAGCAGGCGCCGCGGCCGCCG

CAGCAGCTGCGGGAGGCGCCGGGCAGGGAGGGCAAGGAGGTTACGGGCGTG

GCGGCTACGGACAAGGTGGCGCGGGCCAGGGTGGGGCTGGCGCAGCGGCGG

CGGCTGCCGCCGCAGGCGGCGCTGGGCAAGGCGGGCAGGGTGGATACGGGC

AGGGCGGATATGGCCAGGGTGGGTACGGTCAAGGAGGTGCCGGACAGGGAG

GAGCAGCGGCAGCAGCTGCGGCGGCGGCAGCTGGAGGGGCAGGGCAGGGCG

GTTATGGTCGTGGAGGGGCCGGACAAGGAGGTGCAGCTGCAGCAGCGGGAG

CTGGGCAGGGTGGATACGGAGGTCAGGGCGCAGGACAGGGTGGTGCTGGCG

CTGCGGCTGCGGCAGCCGCGGCGGGAGGCGCGGGGCAGGGAGGCCAAGGCG

GGTATGGAAGAGGAGGCTATGGGCAAGGCGGTGCGGGACAAGGTGGGGCCG

GTGCTGCGGCGGCTGCAGCGGCTGCTGGAGGTGCTGGCCAAGGAGGGCAAG

GCGGATACGGTCAGGGTGGAAATGGGCAAGGCGGAGCGGGCCAGGGCGGCG

CGGCCGCTGCCGCGGCCGCAGCAGGCGGCGCGGGCCAAGGCGGGTACGGTC

GTGGCGGGGCCGGACAAGGCGGAGCTGCCGCCGCCGCTGCGGCCGCGGCTG

GGGCTGGGCAGGGTGGATATGGAGGGCAAGGCGCGGGACAGGGTGGTGCAG

GCGCGGCTGCCGCTGCTGCTGCGGCGGGCGGAGCAGGGCAGGGCGGCCAGG

GTGGATACGGGAGAGGTGGCTACGGTCAAGGAGGGGCCGGCCAAGGTGGTG

CGGGAGCAGCCGCTGCAGCTGCGGCCGCTGGTGGGGCTTCGCAGGGTGGCCA

GGGTGGTTACGGGCAAGGTGATTACGGGCAGGGAGGAGCGGGACAAGGTGG

GGCCGCCGCTGCGGCTGCGGCAGCAGGTGGCGCTGGACAAGGAGGGTATGG

ACGAGGAGGGGCAGGCCAAGGCGGGGCTGCCGCTGCTGCAGGGGCAGGTCA

AGGCGGGTACGGCGGACAAGGTGCAGGGCAAGGAGGTGCTGGAGCTGCGGC

TGCAGCAGCAGCTGCGGGTGGGGCAGGTAGAGGAGGCCAAGGAGGCTACGG

AAGAGGCGGATACGGTCAGGGCGGTGCGGGCCAGGGTGGTGCTGGTGCTGC

CGCAGCCGCCGCTGCTGCCGGTGGGGCTGGACAGGGTGGACAAGGTGGGTA

CGGACAAGGTGGCTACGGGCAGGGAGGCGCGGGTCAAGGAGGGGCGGCGGC

AGCCGCTGCAGCTGCCGCTGGTGGCGCCGGACAAGGAGGCTATGGCCGTGGT

GGGGCGGGGCAAGGCGGCGCGGCGGCCGCGGCGGGTGCGGGGCAGGGCGGC

TATGGAGGTCAAGGAGCCGGCCAAGGTGGGGCAGGCGCAGCAGCCGCTGCG

GCCGCTGCCGGAGGTGCTGGGCGAGGAGGGCAGGGTGGGTATGGTAGAGGC

GGGTACGGGCAGGGAGGCGCCGGACAGGGAGGGGCTGGTGCCGCTGCGGCT

GCGGCAGCCGCCGGAGGAGCGGGTCAGGGTGGCCAGGGCGGCTACGGTCAG

GGCGGGTATGGGCAGGGCGGGGCTGGACAAGGTGGCGCCGCGGCAGCTGCA

GCAGCTGCGGTGGGAGGGGCGGGTCAGGGTGGATACGGCCGAGGTGGGGCA

GGTCAAGGCGGTGCTGCGGCCGCGGCGGCCGCTGCAGCCGCTGGTTCTGGTC

AGGGAGGTTATGGAGGCCAAGGAGCAGGCCAAGGCGGCGCCGGGGCCGCTG

CCGCCGCCGCTGCTGCTGGCGGAGCAGGGCAGGGTGGACAGGGCGGTTACG

GTGGTGGAGGATACGGACAAGGAGGTGCCGGCCAAGGCGGAGCTGGAGCTG

CTGCGGCAGCAGCAGCCGCGGGAGGCGCTGGACAGGGTGGGCAGGGCGGCT

ACGGTCAGGGTGGATACGGTCAGGGCGGGGCGGGGCAGGGTGGAGCCGCAG

CGGCTGCAGCTGCGGCCGCGGGCGGAGCAGGGCAGGGAGGCTATGGTCGGG

GTGGAGCCGGACAAGGAGGCGCGGCGGCCGCTACCGGTGCTGGACAAGGTG

GCTATGGCGGTCAAGGAGCAGGACAAGGCGGAGCGGGCGCCGCTGCTGCAG

CGGCAGCTGCTGGTGGGGCCGGTCAGGGTGGCCAAGGAGGATACGGCAGGG

GTGGCTACGGCCAAGGCGGTGCAGGTCAAGGAGGTGCTGGGGCCGCTGCCG

CCGCCGCAGCCGCCGGAGGAGCTGGGCAAGGCGGACAAGGAGGGTATGGCC

AAGGTGGGTACGGGCAGGGGGGTGCCGGACAAGGCGGAGCGGCCGCAGCGG

CTGCAGCCGCCGGTGGTGCTGGCCAGGGTGGCTATGGGAGAGGCGGGGCCG

GTCAAGGAGGCGCCGCGGCAGCTGCGGCGGCCGCGGCAGGGGCGGGGCAGG

GCGGTTATGGAGGACAAGGCGCGGGTCAGGGTGGAGCGGGTGCGGCGGCGG

CAGCGGCAGCTGCGGGCGGCGCTGGGCAAGGAGGTCAAGGCGGGTACGGTC

GAGGTGGCTACGGACAGGGAGGGGCTGGACAAGGTGGGGCGGGTGCCGCAG

CGGCGGGAGGAGCCGGTCAGGGTGGGCAAGGCGGCTACGGTCAGGGTGGAT

ATGGCCAAGGTGGTGCAGGGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCCG

CAGCCGCAGGCGGCGCGGGCCAAGGAGGATACGGCGGATATGGTCAGCAAG

GCGGAGCTGGAGCGGCTGCTGCAGCGGCTTCAGGACCGGGCCAGATCTACTA

CGGTCCGCAGTCCGTTGCCGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCAG

CACCGGCCACTTCTGCACGCATTTCGTCGCACGCGTCCGCATTGCTTAGTAAT

GGGCCGACAAATCCTGCGTCCATCTCCAACGTAATCAGTAACGCAGTAAGCC

AAATTTCTAGCTCTAACCCAGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGCC
CTACTTGAGCTCGTAACGGCGCTATTGACTATTATCGGCTCATCAAATATTGG
TAGTGTGAACTATGACAGTTCGGGCCAGTACGCACAAGTCGTTACTCAGTCA
GTTCAAAACGCCTTCGCC

Amino Acid Sequence:

GGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAGGAGQGGQGG
YGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAA
AGSGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQ
GGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGYGQGGAGQGGAAAAAAA
AAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAA
AAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYG
QGGNGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGA
GQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGG
AGAAAAAAAAGGASQGGQGGYGQGDYGQGGAGQGGAAAAAAAAGGAGQGG
YGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGRGGQG
GYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQ
GGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGG
AGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQ
GGQGGYGQGGYGQGGAGQGGAAAAAAAAVGGAGQGGYGRGGAGQGGAAA
AAAAAAAGSGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGGGGYG
QGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAA
AAAGGAGQGGYGRGGAGQGGAAAATGAGQGGYGGQGAGQGGAGAAAAAAA
AGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQ

GGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAG

QGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGA

GAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAAGGAGQGGYG

GYGQQGGAGAAAAAASGPGQIYYGPQSVAAPAAAAASALAAPATSARISSHAS

ALLSNGPTNPASISNVISNAVSQISSSNPGASACDVLVQALLELVTALLTIIGSSNIG

SVNYDSSGQYAQVVTQSVQNAFA

A.1.10. 40B LHF1:

Synthetic DNA Sequence:

GGCGGCGCCGGGCAGGGTGGACAGGGTGGCTACGGAAGAGGAGGCTA

CGGTCAAGGAGGCGCAGGCCAAGGCGGGGCAGGAGCTGCTGCTGCCGCGGC

GGCCGCGGGTGGTGCAGGGCAAGGAGGACAGGGAGGGTACGGCCAAGGAGG

ATATGGGCAAGGCGGTTACGGTCAGGGTGGAGCTGGACAAGGCGGAGCGGC

TGCCGCTGCTGCAGCTGCTGGCGGCGCGGGACAAGGTGGTTACGGGAGGGGA

GGTGCCGGACAGGGTGGCGCTGCAGCCGCCGCTGGTGCTGGCCAAGGAGGTT

ACGGTGGCCAAGGTGCTGGGCAGGGTGGGGCTGGTGCGGCTGCTGCGGCGGC

AGCTGCGGGAGGAGCAGGGCAGGGCGGACAAGGTGGCTATGGCAGGGGCGG

TTATGGACAGGGCGGCGCAGGACAAGGAGGAGCGGGCGCGGCGGCTGCAGC

AGCCGCTGCCGGAGGCGCCGGTCAGGGTGGCCAGGGCGGGTACGGCCAAGG

CGGATACGGTCAAGGCGGGGCAGGACAGGGTGGCGCTGCGGCTGCGGCAGC

AGCAGCGGCAGGAGGGGCGGGCCAAGGCGGATATGGGCGGGGCGGAGCGG

GACAAGGCGGAGCGGCAGCTGCAGCGGCCGCAGCGGCGGGTTCAGGCCAGG

GAGGTTACGGCGGCCAAGGTGCCGGACAGGGAGGAGCAGGCGCCGCGGCCG

CCGCAGCAGCTGCGGGAGGCGCCGGGCAGGGAGGGCAAGGAGGTTACGGGC

GTGGCGGCTACGGACAAGGTGGCGCGGGCCAGGGTGGGGCTGGCGCAGCGG

CGGCGGCTGCCGCCGCAGGCGGCGCTGGGCAAGGCGGGCAGGGTGGATACG

GGCAGGGCGGATATGGCCAGGGTGGGTACGGTCAAGGAGGTGCCGGACAGG

GAGGAGCAGCGGCAGCAGCTGCGGCGGCGGCAGCTGGAGGGGCAGGGCAGG

GCGGTTATGGTCGTGGAGGGGCCGGACAAGGAGGTGCAGCTGCAGCAGCGG

GAGCTGGGCAGGGTGGATACGGAGGTCAGGGCGCAGGACAGGGTGGTGCTG

GCGCTGCGGCTGCGGCAGCCGCGGCGGGAGGCGCGGGGCAGGGAGGCCAAG

GCGGGTATGGAAGAGGAGGCTATGGGCAAGGCGGTGCGGGACAAGGTGGGG

CCGGTGCTGCGGCGGCTGCAGCGGCTGCTGGAGGTGCTGGCCAAGGAGGGC

AAGGCGGATACGGTCAGGGTGGAAATGGGCAAGGCGGAGCGGGCCAGGGCG

GCGCGGCCGCTGCCGCGGCCGCAGCAGGCGGCGCGGGCCAAGGCGGGTACG

GTCGTGGCGGGGCCGGACAAGGCGGAGCTGCCGCCGCCGCTGCGGCCGCGG

CTGGGGCTGGGCAGGGTGGATATGGAGGGCAAGGCGCGGGACAGGGTGGTG

CAGGCGCGGCTGCCGCTGCTGCTGCGGCGGGCGGAGCAGGGCAGGGCGGCC

AGGGTGGATACGGGAGAGGTGGCTACGGTCAAGGAGGGGCCGGCCAAGGTG

GTGCGGGAGCAGCCGCTGCAGCTGCGGCCGCTGGTGGGGCTTCGCAGGGTGG

CCAGGGTGGTTACGGGCAAGGTGATTACGGGCAGGGAGGAGCGGGACAAGG

TGGGGCCGCCGCTGCGGCTGCGGCAGCAGGTGGCGCTGGACAAGGAGGGTA

TGGACGAGGAGGGGCAGGCCAAGGCGGGGCTGCCGCTGCTGCAGGGGCAGG

TCAAGGCGGGTACGGCGGACAAGGTGCAGGGCAAGGAGGTGCTGGAGCTGC

GGCTGCAGCAGCAGCTGCGGGTGGGGCAGGTAGAGGAGGCCAAGGAGGCTA

CGGAAGAGGCGGATACGGTCAGGGCGGTGCGGGCCAGGGTGGTGCTGGTGC

TGCCGCAGCCGCCGCTGCTGCCGGTGGGGCTGGACAGGGTGGACAAGGTGG

GTACGGACAAGGTGGCTACGGGCAGGGAGGCGCGGGTCAAGGAGGGGCGGC

GGCAGCCGCTGCAGCTGCCGCTGGTGGCGCCGGACAAGGAGGCTATGGCCGT

GGTGGGGCGGGGCAAGGCGGCGCGGCGGCCGCGGCGGGTGCGGGGCAGGGC

GGCTATGGAGGTCAAGGAGCCGGCCAAGGTGGGGCAGGCGCAGCAGCCGCT

GCGGCCGCTGCCGGAGGTGCTGGGCGAGGAGGGCAGGGTGGGTATGGTAGA

GGCGGGTACGGGCAGGGAGGCGCCGGACAGGGAGGGGCTGGTGCCGCTGCG

GCTGCGGCAGCCGCCGGAGGAGCGGGTCAGGGTGGCCAGGGCGGCTACGGT

CAGGGCGGGTATGGGCAGGGCGGGGCTGGACAAGGTGGCGCCGCGGCAGCT

GCAGCAGCTGCGGTGGGAGGGGCGGGTCAGGGTGGATACGGCCGAGGTGGG

GCAGGTCAAGGCGGTGCTGCGGCCGCGGCGGCCGCTGCAGCCGCTGGTTCTG

GTCAGGGAGGTTATGGAGGCCAAGGAGCAGGCCAAGGCGGCGCCGGGGCCG

CTGCCGCCGCCGCTGCTGCTGGCGGAGCAGGGCAGGGTGGACAGGGCGGTTA

CGGTGGTGGAGGATACGGACAAGGAGGTGCCGGCCAAGGCGGAGCTGGAGC

TGCTGCGGCAGCAGCAGCCGCGGGAGGCGCTGGACAGGGTGGGCAGGGCGG

CTACGGTCAGGGTGGATACGGTCAGGGCGGGGCGGGGCAGGGTGGAGCCGC

AGCGGCTGCAGCTGCGGCCGCGGGCGGAGCAGGGCAGGGAGGCTATGGTCG

GGGTGGAGCCGGACAAGGAGGCGCGGCGGCCGCTACCGGTGCTGGACAAGG

TGGCTATGGCGGTCAAGGAGCAGGACAAGGCGGAGCGGGCGCCGCTGCTGC

AGCGGCAGCTGCTGGTGGGGCCGGTCAGGGTGGCCAAGGAGGATACGGCAG

GGGTGGCTACGGCCAAGGCGGTGCAGGTCAAGGAGGTGCTGGGGCCGCTGC

CGCCGCCGCAGCCGCCGGAGGAGCTGGGCAAGGCGGACAAGGAGGGTATGG

CCAAGGTGGGTACGGGCAGGGGGGTGCCGGACAAGGCGGAGCGGCCGCAGC

GGCTGCAGCCGCCGGTGGTGCTGGCCAGGGTGGCTATGGGAGAGGCGGGGC

CGGTCAAGGAGGCGCCGCGGCAGCTGCGGCGGCCGCGGCAGGGGCGGGGCA

GGGCGGTTATGGAGGACAAGGCGCGGGTCAGGGTGGAGCGGGTGCGGCGGC

GGCAGCGGCAGCTGCGGGCGGCGCTGGGCAAGGAGGTCAAGGCGGGTACGG

TCGAGGTGGCTACGGACAGGGAGGGGCTGGACAAGGTGGGGCGGGTGCCGC

AGCGGCGGGAGGAGCCGGTCAGGGTGGGCAAGGCGGCTACGGTCAGGGTGG

ATATGGCCAAGGTGGTGCAGGGCAGGGTGGGGCTGCAGCTGCTGCCGCAGCC

GCAGCCGCAGGCGGCGCGGGCCAAGGAGGATACGGCGGATATGGTCAGCAA

GGCGGAGCTGGAGCGGCTGCTGCAGCGGCTTCAGGACCGGGCCAGATCTACT

ACGGTCCGCAGTCCGTTGCCGCCCCAGCTGCTGCTGCGGCGTCTGCGTTGGCA

GCACCGGCCACTTCTGCACGCATTTCGTCGCACGCGTCCGCATTGCTTAGTAA

TGGGCCGACAAATCCTGCGTCCATCTCCAACGTAATCAGTAACGCAGTAAGC

CAAATTTCTAGCTCTAACCCAGGCGCTTCTGCCTGTGATGTGCTAGTTCAGGC

CCTACTTGAGCTCGTAACGGCGCTATTGACTATTATCGGCTCATCAAATATTG

GTAGTGTGAACTATGACAGTTCGGGCCAGTACGCACAAGTCGTTACTCAGTC

AGTTCAAAACGCCTTCGCC

Amino Acid Sequence:

GGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGG

YGQGGYGQGGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAA

AGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAG

QGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGG

AGQGGYGRGGAGQGGAAAAAAAAAGSGQGGYGGQGAGQGGAGAAAAAAAA

GGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQG

GYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAG

AGQGGYGGQGAGQGGAGAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQG

GAGAAAAAAAGGAGQGGQGGYGQGGNGQGGAGQGGAAAAAAAGGAGQ

GGYGRGGAGQGGAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAGG

AGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAGGASQGGQGGYGQGDYG

QGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQG

AGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAAAAA

GGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQ

GGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYG

QGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAA

AAVGGAGQGGYGRGGAGQGGAAAAAAAAAAGSGQGGYGGQGAGQGGAGAA

AAAAAAGGAGQGGQGGYGGGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQ

GGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAATGA

GQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGG

AGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAGGAGQG

GYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGA

GQGGQGGYGRGGYGQGGAGQGGAGAAAAGGAGQGGQGGYGQGGYGQGGA

GQGGAAAAAAAAAAGGAGQGGYGGYGQQGGAGAAAAAASGPGQIYYGPQSV

AAPAAAAASALAAPATSARISSHASALLSNGPTNPASISNVISNAVSQISSSNPGAS

ACDVLVQALLELVTALLTIIGSSNIGSVNYDSSGQYAQVVTQSVQNAFA

A.1.11. *Nephila clavipes* Flagelliform Silk (NCF1), Genbank Ascension AF027973

Synthetic DNA Sequence:

GGCCCTGGTGGGGTTGGCCCAGGAGGCAGCGGACCAGGAGGGTATGG
GCCAGGTGGTGCCGGTCCAGGCGGCTATGGTCCAGGAGGTTCGGGACCTGGA
GGGTACGGTCCTGGAGGTAGCGGGCCGGGAGGATATGGACCTGGAGGCTCTG
GACCCGGTGGGTACGGACCGGGAGGTTCAGGCCCCGGTGGCTATGGGCCGG
GTGGGAGCGGTCCCGGAGGTTATGGTCCGGGTGGATACGGTCCGGGAGGGTC
AGGCCCAGGTGGATATGGGCCCGGCGGCACGGGCCCCGGAGGGTCGGGACC
AGGTGGGTATGGCCCCGGCGGATCGGGTCCTGGCGGCTCGGGACCCGGCGGG
TACGGCCCGGGCGGAAGTGGCCCTGGAGGGTTCGGCCCTGGTGGCTCGGGCC
CTGGCGGATACGGTCCAGGCGGCTCAGGACCAGGAGGAGCCGGTCCTGGTG
GAGTGGGACCTGGAGGCTTTGGACCGGGTGGCGCAGGACCTGGTGGTGCTGG
CCCCGGTGGCGCTGGGCCTGGCGGGGCGGGTCCTGGTGGGGCAGGCCCCGGC
GGGGCCGGACCGGGCGGAGCAGGTCCGGGTGGGGCTGGTCCAGGTGGTGCG
GGACCAGGCGGAGCGGGACCCGGAGGGGCCGGCGGCGCAGGCGGTGCGGGA
GGCGCTGGAGGATCGGGTGGTGCAGGAGGGTCTGGAGGGACTACTATTATTG
AAGATCTGGACATTACTATTGATGGGGCGGATGGCCCAATCACTATCTCAGA
AGAGCTTACCATCTCCGGCGCGGGTGGCTCGGGGCCAGGAGGTGCAGGTCCG
GGAGGGGTCGGTCCAGGAGGTAGTGGACCTGGCGGGGTAGGGCCGGGAGGT
AGCGGCCCAGGAGGAGTGGGTCCTGGAGGTTCTGGTCCCGGCGGAGTCGGTC
CCGGAGGGGCCGGTGGGCCCTATGGGCCCGGAGGTTCGGGGCCCGGCGGGG
CGGGAGGTGCCGGCGGTCCGGGTGGAGCTTATGGGCCAGGCGGTAGCTACGG

GCCAGGTGGTTCGGGCGGCCCTGGAGGCGCCGGTGGTCCATACGGGCCTGGC

GGCGAAGGCCCCGGTGGCGCAGGCGGTCCATATGGCCCTGGTGGCGCTGGTG

GTCCCTACGGACCCGGTGGTGCAGGTGGGCCATACGGCCCGGGTGGAGAAG

GAGGTCCTTACGGTCCAGGAGGTTCCTATGGTCCAGGTGGAGCCGGCGGGCC

TTATGGCCCCGGAGGCCCCTATGGCCCGGGCGGGGAAGGTCCTGGCGGGGCT

GGTGGACCATATGGACCCGGCGGTGTAGGCCCTGGAGGTTCCGGACCAGGTG

GTTATGGGCCTGGAGGGTCGGGTCCTGGAGGATACGGACCGGGCGGTGCGGG

CCCAGGTGGGTATGGTCCGGGTGGGAGCGGTCCCGGCGGGTACGGACCTGGT

GGCTCGGGGCCTGGCGGTTACGGCCCTGGCGGAAGCGGTCCGGGAGGGTAC

GGGCCCGGCGGATCAGGTCCAGGCGGCTATGGCAGTGGCGGAGCGGGTCCG

GGCGGATATGGGCCAGGAGGATCGGGACCTGGAGGGTACGGTCCTGGTGGCT

CAGGCCCTGGCGGCTATGGGCCAGGAGGCACGGGGCCTGGAGGGACCGGCC

CGGGTGGTAGTGGACCTGGAGGGTATGGTCCAGGAGGAAGTGGCCCAGGAG

GCTCGGGCCCCGGCGGCTCGGGACCTGGCGGATACGGACCCTCAGGCTCTGG

TCCTGGAGGCTACGGTCCGAGTGGATCGGGCCCTGGAGGTTACGGGCCTGGT

GGATCAGGACCTGGTGGTTACGGACCGGGAGGAAGCGGTGCGGGAGGGACG

GGGCCGGGTGGAGCAGGAGGCGCTGGTGGTGCCGGCGGCAGCGGAGGGGCA

GGTGGTTCTGGTGGCGCAGGCGGTTCGGGTGGCGCCGGAGGATCAGGCGGTG

TTGGTGGGTCCGGTGGAACGACAATCACGGAGGACTTGGATATCACGATCGA

TGGTGCCGACGGTCCCATTACCATCAGTGAGGAACTGACAATCAGTGGTGCT

GGAGGGTCTGGGCCGGGCGGCGCAGGACCCGGCGGCGTCGGACCTGGAGGG

AGTGGACCGGGCGGAGTAGGCCCGGGCGTCAGTGGACCTGGCGGTGTAGGG

CCTGGTGGGTCGGGGCCGGGAGGGGTCGGTTCTGGCGGAAGCGGGCCAGGA

GGCGTGGGACCGGGAGGCTATGGGCCAGGTGGCTCCGGTAGCGGTGGCGTTG
GACCCGGAGGTTACGGTCCGGGCGGGTCGGGTGGGTTTTACGGGCCCGGCGG
TTCTGAAGGGCCATACGGACCCAGCGGAACGTACGGCTCTGGAGGCGGCTAT
GGTCCCGGTGGGGCAGGAGGTCCCTACGGTCCCGGGAGCCCTGGCGGGGCAT
ACGGTCCTGGTTCTCCCGGCGGTGCTTATTATCCCAGTTCTCGGGTGCCGGAC
ATGGTGAATGGTATCATGTCCGCTATGCAAGGTAGTGGTTTTAACTACCAGAT
GTTCGGCAACATGCTATCGCAGTATAGCTCCGGGAGCGGCACATGCAATCCT
AATAATGTCAATGTGCTGATGGACGCTCTACTCGCCGCCTTACATTGCCTATC
AAATCATGGCTCATCCTCATTTGCGCCTAGCCCCACTCCCGCTGCGATGTCGG
CCTACAGCAATAGCGTAGGACGGATGTTCGCCTAT

Amino Acid Sequence:

GPGGVGPGGSGPGGYGPGGAGPGGYGPGGSGPGGYGPGGSGPGGYGPG
GSGPGGYGPGGSGPGGYGPGGSGPGGYGPGGYGPGGSGPGGYGPGGTGPGGSG
PGGYGPGGSGPGGSGPGGYGPGGSGPGGFGPGGSGPGGYGPGGSGPGGAGPGG
VGPGGFGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAG
PGGAGPGGAGGAGGAGGAGGSGGAGGSGGTTIIEDLDITIDGADGPITISEELTIS
GAGGSGPGGAGPGGVGPGGSGPGGVGPGGSGPGGVGPGGSGPGGVGPGGAGGP
YGPGGSGPGGAGGAGGPGGAYGPGGSYGPGGSGGPGGAGGPYGPGGEGPGGA
GGPYGPGGAGGPYGPGGAGGPYGPGGEGGPYGPGGSYGPGGAGGPYGPGGPY
GPGGEGPGGAGGPYGPGGVGPGGSGPGGYGPGGSGPGGYGPGGAGPGGYGPG
GSGPGGYGPGGSGPGGYGPGGSGPGGYGPGGSGPGGYGSGGAGPGGYGPGGSG
PGGYGPGGSGPGGYGPGGTGPGGTGPGGSGPGGYGPGGSGPGGSGPGGSGPGG
YGPSGSGPGGYGPSGSGPGGYGPGGSGPGGYGPGGSGAGGTGPGGAGGAGGAG

GSGGAGGSGGAGGSGGAGGSGGVGGSGGTTITEDLDITIDGADGPITISEELTISG
AGGSGPGGAGPGGVGPGGSGPGGVGPGVSGPGGVGPGGSGPGGVGSGGSGPGG
VGPGGYGPGGSGSGGVGPGGYGPGGSGGFYGPGGSEGPYGPSGTYGSGGGYGP
GGAGGPYGPGSPGGAYGPGSPGGAYYPSSRVPDMVNGIMSAMQGSGFNYQMF
GNMLSQYSSGSGTCNPNNVNVLMDALLAALHCLSNHGSSSFAPSPTPAAMSAYS
NSVGRMFAY

A.1.12. 1B NCF1:

Synthetic DNA Sequence:

TCCGGTGGAACGACAATCACGGAGGACTTGGATATCACGATCGATGGT
GCCGACGGTCCCATTACCATCAGTGAGGAACTGACAATCAGTGGTGCTGGAG
GGTCTGGGCCGGGCGGCGCAGGACCCGGCGGCGTCGGACCTGGAGGGAGTG
GACCGGGCGGAGTAGGCCCGGGCGTCAGTGGACCTGGCGGTGTAGGGCCTG
GTGGGTCGGGGCCGGGAGGGGTCGGTTCTGGCGGAAGCGGGCCAGGAGGCG
TGGGACCGGGAGGCTATGGGCCAGGTGGCTCCGGTAGCGGTGGCGTTGGACC
CGGAGGTTACGGTCCGGGCGGGTCGGGTGGGTTTTACGGGCCCGGCGGTTCT
GAAGGGCCATACGGACCCAGCGGAACGTACGGCTCTGGAGGCGGCTATGGT
CCCGGTGGGGCAGGAGGTCCCTACGGTCCCGGGAGCCCTGGCGGGGCATACG
GTCCTGGTTCTCCCGGCGGTGCTTATTATCCCAGTTCTCGGGTGCCGGACATG
GTGAATGGTATCATGTCCGCTATGCAAGGTAGTGGTTTTAACTACCAGATGTT
CGGCAACATGCTATCGCAGTATAGCTCCGGGAGCGGCACATGCAATCCTAAT
AATGTCAATGTGCTGATGGACGCTCTACTCGCCGCCTTACATTGCCTATCAAA

TCATGGCTCATCCTCATTTGCGCCTAGCCCCACTCCCGCTGCGATGTCGGCCT
ACAGCAATAGCGTAGGACGGATGTTCGCCTAT

Amino Acid Sequence:

SGGTTITEDLDITIDGADGPITISEELTISGAGGSGPGGAGPGGVGPGGSGP
GGVGPGVSGPGGVGPGGSGPGGVGSGGSGPGGVGPGGYGPGGSGSGGVGPGGY
GPGGSGGFYGPGGSEGPYGPSGTYGSGGGYGPGGAGGPYGPGSPGGAYGPGSPG
GAYYPSSRVPDMVNGIMSAMQGSGFNYQMFGNMLSQYSSGSGTCNPNNVNVL
MDALLAALHCLSNHGSSSFAPSPTPAAMSAYSNSVGRMFAY

A.1.13. 2B NCF1:

Synthetic DNA Sequence:

ACTACTATTATTGAAGATCTGGACATTACTATTGATGGGGCGGATGGC
CCAATCACTATCTCAGAAGAGCTTACCATCTCCGGCGCGGGTGGCTCGGGGC
CAGGAGGTGCAGGTCCGGGAGGGGTCGGTCCAGGAGGTAGTGGACCTGGCG
GGGTAGGGCCGGGAGGTAGCGGCCCAGGAGGAGTGGGTCCTGGAGGTTCTG
GTCCCGGCGGAGTCGGTCCCGGAGGGGCCGGTGGGCCCTATGGGCCCGGAG
GTTCGGGGCCCGGCGGGGCGGGAGGTGCCGGCGGTCCGGGTGGAGCTTATGG
GCCAGGCGGTAGCTACGGGCCAGGTGGTTCGGGCGGCCCTGGAGGCGCCGGT
GGTCCATACGGGCCTGGCGGCGAAGGCCCCGGTGGCGCAGGCGGTCCATATG
GCCCTGGTGGCGCTGGTGGTCCCTACGGACCCGGTGGTGCAGGTGGGCCATA
CGGCCCGGGTGGAGAAGGAGGTCCTTACGGTCCAGGAGGTTCCTATGGTCCA
GGTGGAGCCGGCGGGCCTTATGGCCCCGGAGGCCCCTATGGCCCGGGCGGGG

AAGGTCCTGGCGGGGCTGGTGGACCATATGGACCCGGCGGTGTAGGCCCTGG

AGGTTCCGGACCAGGTGGTTATGGGCCTGGAGGGTCGGGTCCTGGAGGATAC

GGACCGGGCGGTGCGGGCCCAGGTGGGTATGGTCCGGGTGGGAGCGGTCCC

GGCGGGTACGGACCTGGTGGCTCGGGGCCTGGCGGTTACGGCCCTGGCGGAA

GCGGTCCGGGAGGGTACGGGCCCGGCGGATCAGGTCCAGGCGGCTATGGCA

GTGGCGGAGCGGGTCCGGGCGGATATGGGCCAGGAGGATCGGGACCTGGAG

GGTACGGTCCTGGTGGCTCAGGCCCTGGCGGCTATGGGCCAGGAGGCACGGG

GCCTGGAGGGACCGGCCCGGGTGGTAGTGGACCTGGAGGGTATGGTCCAGG

AGGAAGTGGCCCAGGAGGCTCGGGCCCCGGCGGCTCGGGACCTGGCGGATA

CGGACCCTCAGGCTCTGGTCCTGGAGGCTACGGTCCGAGTGGATCGGGCCCT

GGAGGTTACGGGCCTGGTGGATCAGGACCTGGTGGTTACGGACCGGGAGGA

AGCGGTGCGGGAGGGACGGGGCCGGGTGGAGCAGGAGGCGCTGGTGGTGCC

GGCGGCAGCGGAGGGGCAGGTGGTTCTGGTGGCGCAGGCGGTTCGGGTGGC

GCCGGAGGATCAGGCGGTGTTGGTGGGTCCGGTGGAACGACAATCACGGAG

GACTTGGATATCACGATCGATGGTGCCGACGGTCCCATTACCATCAGTGAGG

AACTGACAATCAGTGGTGCTGGAGGGTCTGGGCCGGGCGGCGCAGGACCCG

GCGGCGTCGGACCTGGAGGGAGTGGACCGGGCGGAGTAGGCCCGGGCGTCA

GTGGACCTGGCGGTGTAGGGCCTGGTGGGTCGGGGCCGGGAGGGGTCGGTTC

TGGCGGAAGCGGGCCAGGAGGCGTGGGACCGGGAGGCTATGGGCCAGGTGG

CTCCGGTAGCGGTGGCGTTGGACCCGGAGGTTACGGTCCGGGCGGGTCGGGT

GGGTTTTACGGGCCCGGCGGTTCTGAAGGGCCATACGGACCCAGCGGAACGT

ACGGCTCTGGAGGCGGCTATGGTCCCGGTGGGGCAGGAGGTCCCTACGGTCC

CGGGAGCCCTGGCGGGGCATACGGTCCTGGTTCTCCCGGCGGTGCTTATTATC

CCAGTTCTCGGGTGCCGGACATGGTGAATGGTATCATGTCCGCTATGCAAGG
TAGTGGTTTTAACTACCAGATGTTCGGCAACATGCTATCGCAGTATAGCTCCG
GGAGCGGCACATGCAATCCTAATAATGTCAATGTGCTGATGGACGCTCTACT
CGCCGCCTTACATTGCCTATCAAATCATGGCTCATCCTCATTTGCGCCTAGCC
CCACTCCCGCTGCGATGTCGGCCTACAGCAATAGCGTAGGACGGATGTTCGC
CTAT

Amino Acid Sequence:

TTIIEDLDITIDGADGPITISEELTISGAGGSGPGGAGPGGVGPGGSGPGGVG
PGGSGPGGVGPGGSGPGGVGPGGAGGPYGPGGSGPGGAGGAGGPGGAYGPGGS
YGPGGSGGPGGAGGPYGPGGEGPGGAGGPYGPGGAGGPYGPGGAGGPYGPGG
EGGPYGPGGSYGPGGAGGPYGPGGPYGPGGEGPGGAGGPYGPGGVGPGGSGPG
GYGPGGSGPGGYGPGGAGPGGYGPGGSGPGGYGPGGSGPGGYGPGGSGPGGYG
PGGSGPGGYGSGGAGPGGYGPGGSGPGGYGPGGSGPGGYGPGGTGPGGTGPGG
SGPGGYGPGGSGPGGSGPGGSGPGGYGPSGSGPGGYGPSGSGPGGYGPGGSGPG
GYGPGGSGAGGTGPGGAGGAGGAGGSGGAGGSGGAGGSGGAGGSGGVGGSG
GTTITEDLDITIDGADGPITISEELTISGAGGSGPGGAGPGGVGPGGSGPGGVGPGV
SGPGGVGPGGSGPGGVGSGGSGPGGVGPGGYGPGGSGSGGVGPGGYGPGGSGG
FYGPGGSEGPYGPSGTYGSGGGYGPGGAGGPYGPGSPGGAYGPGSPGGAYYPSS
RVPDMVNGIMSAMQGSGFNYQMFGNMLSQYSSGSGTCNPNNVNVLMDALLAA
LHCLSNHGSSSFAPSPTPAAMSAYSNSVGRMFAY

A.1.14. *Bombyx mori* Heavy Chain Fibroin 3' partial (BMF1), Genbank Ascension S74439

Synthetic DNA Sequence:

GCGGGAACGGGGTCATCAGGTTTTGGCCCTTATGTAGCGAACGGTGGC
TATTCTGGCTATGAGTACGCTTGGTCTAGTGAAAGTGATTTTGGTACCGGTAG
TGGGGCAGGCGCAGGCAGCGGCGCTGGGGCCGGTAGTGGGGCGGGAGCTGG
ATATGGTGCCGGTGTTGGCGCAGGCTATGGGGCCGGATATGGTGCAGGTGCG
GGAGCAGGGTACGGGGCAGGCGCTGGAAGTGGGGTCGCTTCGGGAGCCGGT
GCCGGCGCGGGGTCTGGAGCCGGCGCGGGAAGCGGGGCGGGTGCCGGGTCA
GGAGCGGGTGCTGGTTCAGGCGCAGGAGCTGGCTCCGGGGCGGGTGCAGGG
TCCGGCGCAGGGGCTGGTTATGGCGCTGGCGCTGGTTATGGCGCTGGCGCAG
GTTATGGGGCGGGTGCTGGAGTCGGGTATGGTGCTGGCGCAGGTGTAGGATA
CGGTGCGGGTGCGGGATATGGGGCGGGAGCCGGTGTGGGGTATGGTGCCGG
GGCAGGGAGTGGCGCGGCCTCAGGCGCCGGGGCCGGCTCAGGAGCAGGCGC
GGGAAGCGGGGCTGGCGCAGGTTCGGGTGCAGGAGCAGGCTCGGGAGCTGG
TGCAGGATCGGGAGCTGGGGCAGGCAGTGGGGCCGGAGCAGGTTCCGGTGC
TGGCGCAGGTAGCGGAGCCGGTGCGGGTTATGGCGCGGGAGCTGGAGTAGG
CTATGGTGCAGGTGCGGGTTCAGGAGCTGCTAGTGGCGCGGGCGCTGGTTCC
GGAGCAGGTGCTGGCTCGGGCGCCGGAGCCGGGAGTGGCGCAGGCGCTGGA
TCAGGTGCCGGAGCAGGATCTGGGGCTGGTGCCGGATCGGGAGCCGGTGCCG
GGTCAGGGGCCGGCGCTGGGTCGGGCGCTGGAAGTGGTGCAGGGGCAGGTT
CAGGCGCAGGAGCGGGTTATGGTGCGGGTGCCGGAGCAGGCGTCGGGTACG
GAGCTGGGGCCGGTGCAGGTTATGGAGCGGGCTACGGTTATGGCGCTGGTGC

CGGGGTGGGTTACGGTGCGGGCGCAGGTTCTGGCGCCGCCAGTGGGGCCGGC

GCAGGCTCAGGTGCAGGAGCCGGTAGTGGGGCGGGCGCCGGGAGTGGGGCC

GGCGCAGGCAGCGGAGCCGGGGCTGGTTCAGGAGCGGGGTCAGGCGCGGGA

GCTGGAAGTGGGGCCGGGGCGGGTTACGGTGCCGGATACGGAGCCGGTGTTG

GGGCAGGGTACGGAGCCGGCGCAGGCGTAGGGTATGGGGCAGGATATGGTG

TAGGAGCTGGAGCAGGTTACGGCGCAGGTGCTGGGTCTGGCGCCGCATCAGG

AGCCGGGGCCGGCAGTGGTGCGGGCGCTGGGTCTGGGGCCGGTGCCGGTTCC

GGGGCGGGAGCAGGCAGCGGTGCAGGGGCGGGATCGGGCGCTGGCTCAGGC

GCCGGAGCTGGGTATGGGGCGGGTGCTGGAAGTGGAGCGGCGTCAGGAGCC

GGAGCAGGTGCCGGGGCTGGTACAGGTAGCAGCGGATTCGGACCATACGTTG

CCAATGGTGGATATTCCCGACGCGAAGGGTACGAATATGCCTGGTCTAGTAA

ATCGGATTTTGAGACAGGATCAGGTGCGGCTTCGGGAGCTGGGGCAGGCGCC

GGGTCAGGGGCCGGTGCAGGTTCGGGCGCGGGTGCTGGATCTGGGGCTGGCG

CAGGCAGTGGTGCGGGGGCGGGAGGTAGCGTTTCTTACGGTGCCGGTAGAGG

GTATGGGCAGGGCGCTGGCTCGGCCGCGAGCTCCGTGTCCAGTGCTTCGAGT

AGAAGTTACGATTACTCGCGCCGTAATGTCAGGAAGAATTGTGGTATCCCAC

GACGCCAGCTCGTAGTAAAATTTCGGGCCCTCCCGTGCGTAAACTGC

Amino Acid Sequence:

AGTGSSGFGPYVANGGYSGYEYAWSSESDFGTGSGAGAGSGAGAGSGA

GAGYGAGVGAGYGAGYGAGAGAGYGAGAGSGVASGAGAGAGSGAGAGSGA

GAGSGAGAGSGAGAGSGAGAGSGAGAGYGAGAGYGAGAGYGAGAGVGYGA

GAGVGYGAGAGYGAGAGVGYGAGAGSGAASGAGAGSGAGAGSGAGAGSGA

GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGYGAGAGVGYGAGAGSGAA

SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA

GSGAGAGSGAGAGYGAGAGAGVGYGAGAGAGYGAGYGYGAGAGVGYGAGA

GSGAASGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGSGAGAGSGAGAG

YGAGYGAGVGAGYGAGAGVGYGAGYGVGAGAGYGAGAGSGAASGAGAGSG

AGAGSGAGAGSGAGAGSGAGAGSGAGSGAGAGYGAGAGSGAASGAGAGAGA

GTGSSGFGPYVANGGYSRREGYEYAWSSKSDFETGSGAASGAGAGAGSGAGAG

SGAGAGSGAGAGSGAGAGGSVSYGAGRGYGQGAGSAASSVSSASSRSYDYSRR

NVRKNCGIPRRQLVVKFRALPCVNC

A.1.15. *Araneus diadematus* ADF1, Genbank Ascension U47853

Synthetic DNA Sequence:

CATGAATCTTCCTATGCTGCTGCAATGGCTGCTTCTACTCGTAATTCTG

ATTTTATCCGTAACATGAGCTACCAGATGGGTCGTCTGCTGAGCAACGCCGGT

GCCATTACCGAATCTACTGCAAGCAGCGCGGCTTCCAGCGCGTCCTCCACCG

TTACCGAGTCTATTCGCACGTATGGCCCGGCTGCGATCTTTTCTGGTGCGGGC

GCTGGCGCAGGCGTGGGTGTAGGTGGTGCCGGTGGTTACGGCCAGGGCTACG

GTGCAGGCGCAGGTGCTGGTGCGGGCGCCGGTGCGGGTGCTGGTGGCGCGGG

TGGCTACGGTCAGGGCTACGGTGCGGGTGCTGCCGCGGCGGCTGGCGCAGGC

GCAGGTGCAGCGGGCGGTTACGGCGGTGGTTCTGGTGCTGGCGCTGGTGGTG

CTGGTGGTTATGGCCAGGGTTACGGTGCAGGTTCTGGCGCGGGTGCGGGCGC

TGCTGCGGCAGCTGGCGCATCCGCTGGTGCTGCTGGCGGCTATGGCGGTGGC

GCAGGTGTTGGTGCAGGTGCGGGCGCGGGTGCGGCTGGTGGCTATGGCCAGA

GCTATGGCAGCGGTGCTGGCGCAGGTGCGGGTGCTGGTGCGGCGGCTGCAGC

TGGCGCTGGCGCACGTGCAGCGGGTGGCTACGGTGGTGGTTACGGCGCAGGC

GCGGGCGCCGGTGCTGGCGCCGCTGCTTCCGCTGGTGCCTCCGGTGGCTACG
GTGGCGGTTACGGCGGTGGCGCGGGTGCAGGCGCCGTAGCTGGTGCGTCCGC
GGGTTCTTACGGCGGTGCGGTTAACCGTCTGTCTAGCGCAGGCGCGGCATCTC
GTGTTTCCAGCAACGTGGCTGCCATCGCGTCTGCGGGTGCGGCTGCGCTGCC
GAACGTAATCTCTAACATTTATTCTGGTGTGCTGTCCTCTGGTGTGTCTTCTTC
TGAGGCGCTGATCCAGGCTCTGCTGGAAGTCATCTCTGCACTGATCCACGTGC
TGGGTTCTGCCTCCATCGGTAACGTGTCTTCCGTTGGCGTTAACAGCGCACTG
AATGCAGTGCAGAACGCCGTCGGCGCGTACGCTGGT

Amino Acid Sequence:

HESSYAAAMAASTRNSDFIRNMSYQMGRLLSNAGAITESTASSAASSASS
TVTESIRTYGPAAIFSGAGAGAGVGVGGAGGYGQGYGAGAGAGAGAGAGG
AGGYGQGYGAGAAAAAGAGAGAAGGYGGGSGAGAGGAGGYGQGYGAGSGA
GAGAAAAAGASAGAAGGYGGGAGVGAGAGAGAAGGYGQSYGSGAGAGAGA
GAAAAAGAGARAAGGYGGGYGAGAGAGAGAAASAGASGGYGGGYGGGAGA
GAVAGASAGSYGGAVNRLSSAGAASRVSSNVAAIASAGAAALPNVISNIYSGVL
SSGVSSSEALIQALLEVISALIHVLGSASIGNVSSVGVNSALNAVQNAVGAYAG

A.1.16. *Araneus diadematus* ADF2, Genbank Ascension U47854

Synthetic DNA Sequence:

GGTAGCCAAGGCGCAGGTGGTGCAGGTCAAGGTGGTTATGGTGCAGG
CGGCGGTGGCGCTGCGGCAGCTGCTGCTGCAGCGGTAGGCGCGGGTGGCGGT
GGTCAGGGCGGCCTGGGTTCCGGCGGTGCGGGCCAGGGTTACGGCGCAGGCC
TGGGCGGTCAAGGTGGCGCATCTGCGGCGGCTGCTGCGGCTGGTGGCCAGGG

CGGTCAGGGTGGCCAAGGTGGCTATGGCGGTCTGGGTTCTCAGGGCGCAGGC
GGTGCTGGTCAGCTGGGCTATGGTGCAGGTCAGGAATCTGCAGCGGCTGCCG
CTGCCGCAGCGGGCGGCGCTGGTGGCGGTGGTCAGGGCGGCCTGGGTGCGG
GTGGCGCTGGCCAAGGTTACGGTGCCGCTGGCCTGGGCGGTCAGGGTGGTGC
GGGCCAGGGCGGCGGCTCTGGCGCGGCGGCTGCGGCCGGTGGTCAAGGTGG
TCAGGGCGGCTATGGTGGCCTGGGCCCGCAAGGCGCGGGTGGTGCGGGCCA
GGGTGGCTACGGTGGTGGTTCCCTGCAATACGGCGGTCAGGGTCAGGCTCAG
GCAGCTGCGGCATCCGCGGCGGCATCCCGCCTGTCTTCCCCATCCGCAGCGG
CACGTGTGTCTTCCGCTGTATCTCTGGTATCCAACGGCGGTCCGACCAGCCCG
GCGGCACTGAGCTCTAGCATTTCCAACGTGGTATCTCAGATCTCTGCAAGCAA
CCCAGGCCTGTCTGGTTGCGATATCCTGGTTCAAGCCCTGCTGGAAATTATCT
CTGCGCTGGTTCACATCCTGGGTTCTGCCAACATCGGCCCGGTTAACTCTAGC
TCCGCCGGTCAGTCCGCATCCATTGTAGGTCAATCCGTATACCGCGCTCTGTC
T

Amino Acid Sequence:

GSQGAGGAGQGGYGAGGGGAAAAAAAAVGAGGGGQGGLGSGGAGQG
YGAGLGGQGGASAAAAAAGGQGGGQGGGQGGYGGLGSQGAGGAGQLGYGAGQ
ESAAAAAAAAGGAGGGGQGGLGAGGAGQGYGAAGLGGQGGAGQGGGSGAA
AAAGGQGGGQGGYGGLGPQGAGGAGQGGYGGGSLQYGGGQGQAQAAAASAAAS
RLSSPSAAARVSSAVSLVSNGGPTSPAALSSSISNVVSQISASNPGLSGCDILVQAL
LEIISALVHILGSANIGPVNSSSAGQSASIVGQSVYRALS

A.1.17. *Araneus diadematus* ADF3, Genbank Ascension U47855

Synthetic DNA Sequence:

GCCCGCGCGGGGTCAGGCCAGCAGGGACCAGGTCAACAGGGCCCGGG
CCAACAAGGCCCGGGTCAACAGGGTCCGTACGGTCCGGGTGCCAGCGCGGC
GGCCGCGGCCGCAGGAGGGTATGGCCCTGGTAGCGGCCAACAGGGTCCGAG
CCAGCAAGGCCCGGGCCAGCAAGGGCCGGGGGGCCAGGGGCCCTACGGCCC
TGGTGCGTCAGCTGCCGCAGCCGCAGCTGGCGGTTATGGCCCGGGGTCAGGT
CAGCAAGGGCCAGGCGGTCAAGGTCCTTACGGGCCAGGCAGTAGTGCGGCA
GCGGCTGCTGCCGGTGGTAACGGCCCGGGGTCGGGCCAGCAAGGGGCGGGA
CAGCAGGGTCCAGGCCAACAAGGCCCCGGTGCGTCCGCAGCGGCGGCGGCC
GCTGGTGGCTATGGCCCGGGTTCAGGCCAGCAGGGCCCGGGGCAGCAGGGC
CCGGGTGGACAGGGTCCGTATGGCCCGGGGGCCAGTGCAGCGGCCGCGGCT
GCTGGGGGCTATGGCCCTGGCTCAGGTCAGGGTCCGGGTCAACAAGGACCCG
GCGGTCAAGGACCGTATGGCCCGGGTGCGTCCGCGGCGGCTGCGGCGGCTGG
AGGCTATGGTCCGGGAAGTGGCCAACAGGGCCCTGGACAGCAGGGTCCGGG
TCAGCAGGGACCCGGTGGACAGGGCCCGTATGGGCCAGGCGCCTCTGCCGCA
GCGGCGGCCGCAGGTGGGTATGGACCGGGGTACGGCCAGCAGGGTCCTGGT
CAGCAGGGACCGGGCGGCCAGGGCCCTTACGGCCCCGGCGCGTCAGCTGCA
AGCGCTGCCTCGGGTGGCTACGGCCCGGGTTCCGGTCAGCAGGGCCCGGGAC
AGCAGGGTCCGGGTCAGCAGGGACCGTATGGTCCGGGAGCTTCTGCTGCTGC
CGCCGCGGCGGGTGGTTATGGACCCGGCAGTGGCCAACAAGGTCCGGGGCA
GCAGGGTCCAGGTCAGCAGGGCCCAGGACAGCAGGGCCCTGGTGGCCAAGG
ACCGTACGGTCCCGGCGCAAGTGCGGCCGCTGCAGCTGCCGGAGGCTACGGT

CCAGGTAGTGGACAGCAAGGACCGGGTCAGCAGGGCCCCGGTCAACAGGGG

CCGGGCCAGCAAGGCCCCGGGCAGCAGGGACCTGGGCAGCAGGGTCCCGGG

CAGCAAGGTCCTGGGCAACAGGGTCCGGGACAGCAAGGCCCTGGCGGCCAG

GGTGCGTATGGGCCTGGTGCATCTGCCGCGGCGGGCGCCGCGGGTGGGTACG

GGCCGGGGAGCGGCCAGCAAGGTCCGGGCCAACAGGGCCCCGGACAACAGG

GTCCTGGCCAGCAAGGACCTGGCCAGCAGGGGCCGGGACAACAAGGGCCCG

GCCAACAAGGCCCAGGGCAACAAGGCCCGTACGGCCCTGGGGCCTCGGCAG

CCGCGGCAGCGGCCGGCGGCTATGGCCCGGGCAGTGGTCAACAAGGTCCAG

GCCAACAGGGCCCAGGGCAGCAGGGTCCGGGGGGGTCAAGGTCCGTACGGAC

CGGGTGCCGCCTCGGCAGCGGTGAGTGTAGGCGGCTACGGACCTCAAAGCTC

CTCTGTGCCAGTCGCCAGTGCGGTGGCTAGCCGTCTGTCTAGCCCCGCCGCCA

GCAGTCGTGTCAGCTCAGCCGTGTCGTCTTTAGTATCATCAGGACCGACTAAA

CACGCAGCCTTGTCAAACACCATTAGCAGCGTTGTCTCTCAGGTGTCAGCGA

GTAACCCGGGGCTGTCGGGTTGCGACGTCCTGGTACAGGCCCTGCTGGAAGT

GGTGAGCGCCCTCGTGTCTATTCTGGGTTCTAGTTCCATTGGCCAGATTAACT

ATGGGGCGAGTGCGCAATACACCCAAATGGTCGGACAATCTGTTGCGCAGGC

ACTGGCG

Amino Acid Sequence:

ARAGSGQQGPGQQGPGQQGPGQQGPYGPGASAAAAAAGGYGPGSGQQ

GPSQQGPGQQGPGGQGPYGPGASAAAAAAGGYGPGSGQQGPGGQGPYGPGSSA

AAAAAGGNGPGSGQQGAGQQGPGQQGPGASAAAAAAGGYGPGSGQQGPGQQ

GPGGQGPYGPGASAAAAAAGGYGPGSGQGPGQQGPGGQGPYGPGASAAAAAA

GGYGPGSGQQGPGQQGPGQQGPGGQGPYGPGASAAAAAAGGYGPGYGQQGPG

QQGPGGQGPYGPGASAASAASGGYGPGSGQQGPGQQGPGQQGPYGPGASAAA
AAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGGQGPYGPGASAAAAAAGGYGP
GSGQQGPGQQGPGQQGPGQQGPGQQGPGQQGPGQQGPGQQGPGQQGPGGQGA
YGPGASAAAGAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQGPGQQGPGQQ
GPGQQGPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGGQGPYGPGAA
SAAVSVGGYGPQSSSVPVASAVASRLSSPAASSRVSSAVSSLVSSGPTKHAALSN
TISSVVSQVSASNPGLSGCDVLVQALLEVVSALVSILGSSSIGQINYGASAQYTQM
VGQSVAQALA

A.1.18. *Araneus diadematus* ADF4, Genbank Ascension U47856

Synthetic DNA Sequence:

GCAGGCTCTAGCGCCGCAGCTGCCGCTGCTGCAAGCGGTAGCGGTGGT
TACGGTCCAGAGAACCAGGGTCCGTCCGGCCCAGTAGCATATGGCCCTGGTG
GTCCAGTCTCTTCCGCTGCTGCCGCAGCTGCTGCGGGCTCCGGTCCAGGTGGC
TACGGTCCGGAAAACCAGGGCCCGTCTGGTCCGGGCGGTTATGGCCCGGGTG
GCTCTGGTAGCTCTGCAGCGGCGGCAGCCGCGGCAGCGTCTGGCCCAGGTGG
TTACGGCCCAGGCTCCCAGGGCCCGTCCGGTCCGGGCGGTAGCGGCGGTTAT
GGTCCTGGTTCCCAGGGTGCAAGCGGCCCTGGTGGTCCGGGCGCATCTGCGG
CAGCCGCCGCAGCAGCGGCTGCGGCAAGCGGTCCGGGTGGCTACGGTCCGG
GCAGCCAGGGTCCGTCTGGTCCTGGCGCCTACGGTCCAGGTGGCCCGGGTTC
CTCCGCTGCAGCCGCGGCTGCGGCTGCGAGCGGTCCTGGTGGCTACGGTCCG
GGTAGCCAGGGTCCTTCCGGTCCAGGTGTGTACGGCCCTGGTGGCCCGGGTT
CTTCTGCTGCTGCAGCGGCGGCTGCTGGCTCTGGTCCGGGCGGTTATGGCCCG
GAAAACCAGGGTCCGTCTGGCCCTGGTGGTTACGGCCCAGGTGGTTCCGGTT

CCTCCGCTGCTGCGGCGGCAGCAGCTGCCAGCGGTCCAGGCGGTTACGGTCC
TGGCTCTCAAGGCCCGTCCGGCCCTGGCGGTTCCGGTGGCTATGGTCCGGGTT
CTCAGGGCGGTTCTGGTCCGGGCGCGAGCGCAGCTGCAGCGGCAGCCGCTGC
ATCTGGTCCTGGCGGTTACGGTCCGGGTAGCCAGGGTCCATCCGGTCCGGGC
TATCAGGGCCCGTCTGGCCCGGGTGCTTATGGTCCATCCCCGAGCGCATCTGC
GTCCGTGGCCGCTTCCGTCTATCTGCGTCTGCAACCGCGTCTGGAAGTTTCCT
CTGCTGTTAGCAGCCTGGTTTCCAGCGGTCCGACTAACGGCGCTGCTGTCTCT
GGCGCCCTGAACTCTCTGGTTTCCCAGATTTCTGCAAGCAACCCTGGTCTGTC
TGGTTGCGACGCGCTGGTGCAGGCTCTGCTGGAACTGGTTTCTGCGCTGGTTG
CAATCCTGAGCAGCGCAAGCATCGGTCAGGTTAACGTTTCTTCTGTCAGCCAG
AGCACCCAAATGATTTCTCAGGCACTGAGC

Amino Acid Sequence:

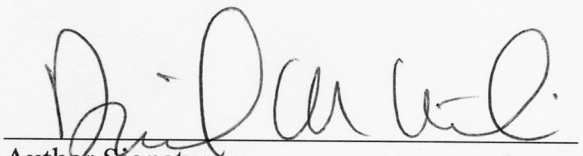
AGSSAAAAAAASGSGGYGPENQGPSGPVAYGPGGPVSSAAAAAAAGSGP
GGYGPENQGPSGPGGYGPGGSGSSAAAAAAAASGPGGYGPGSQGPSGPGGSGG
YGPGSQGASGPGGPGASAAAAAAAAAASGPGGYGPGSQGPSGPGAYGPGGPGS
SAAAAAAAASGPGGYGPGSQGPSGPGVYGPGGPGSSAAAAAAAGSGPGGYGPE
NQGPSGPGGYGPGGSGSSAAAAAAAASGPGGYGPGSQGPSGPGGSGGYGPGSQ
GGSGPGASAAAAAAAASGPGGYGPGSQGPSGPGYQGPSGPGAYGPSPSASASVA
ASVYLRLQPRLEVSSAVSSLVSSGPTNGAAVSGALNSLVSQISASNPGLSGCDAL
VQALLELVSALVAILSSASIGQVNVSSVSQSTQMISQALS

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

5/6/2010

Date