ENGINEERING EXTRACELLULAR SECRETION PATHWAYS IN *ESCHERICHIA COLI*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Charles Henry Haitjema

August 2012

ENGINEERING EXTRACELLULAR SECRETION PATHWAYS IN *ESCHERICHIA COLI*

Charles Henry Haitjema, Ph. D.

Cornell University 2012

Extracellular secretion is highly desirable in preparative protein production. The bacterium *Escherichia coli* is a commonly used for both laboratory- and industrial- scale biosynthesis of proteins, but it lacks many of the pathways for exporting proteins out of cells. This lack of a dedicated extracellular secretion system represents a major bottleneck across many biotechnology disciplines, in particular the bioprocessing of plant biomass where extracellular secretion of cellulase is required. Furthermore, the study and engineering of extracellular secretion systems is limited due to a lack of high-throughput screen to identify rare genetic conditions that affect secretion activity. Recently, it was discovered in *E. coli* that the YebF protein is secreted efficiently into the supernatant when over expressed, and YebF has been employed to carry heterologous proteins into the supernatant via C-terminal genetic fusions. Here, we harness the YebF pathway to simultaneously co-secrete active cellulases into the culture medium, which enabled non-cellulolytic *E. coli* cells to utilize and convert cellulose to bioenergy products. We also developed a universal approach to study and engineer YebF and other extracellular secretion pathways. This high-throughput screening platform was used to screen a genome-wide transposon insertion library for the isolation of gene deletions that upregulate the secretion of YebF and YebF fusions. We also developed an alternative strategy for engineering extracellular secretion systems by way of a genetic selection where the non-secretory phenotype is lethal. Finally, we describe some of the physiological consequences to the bacterial host caused by heterologous protein secretion, in particular an envelope stress response that triggers CRISPR RNA-mediated DNA silencing.

BIOGRAPHICAL SKETCH

Charles Haitjema was born to Dutch parents in Bloomington, Indiana. He received a Bachelor of Arts in Spanish and Bachelor of Science in Microbiology from Indiana University in 2007, and immediately enrolled in Cornell University to pursue a Ph.D. in Microbiology. He accepted a postdoctoral fellowship position at the University of California, Santa Barbara.

This work would not have been possible without the love and support of my parents, Bieneke and Henk, my sister, Coraline, and my brother, Mart. I also thank past and present advisors for guidance and inspiration, particularly my undergraduate mentor Dr. Clay Fuqua and my graduate advisor Dr. Matthew DeLisa.

ACKNOWLEDGMENTS

This work was performed with financial support from:

Department of Energy Great Lakes Bioenergy Research Center (GLBRC)

In addition technical support was provided by:

Dr. Adam Fisher

Jason Boock

Dr. Brian King

Dr. Jan Kostecki

Dr. Ritsdeliz Perez-Rodriguez

Dr. Dujduan Waraho-Zhmayev

Dr. Sydnor Withers

Miguel Dominguez

Dr. Jeffrey Gardner

Dr. David Keating

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Escherichia coli (*E. coli*)
General secretory (Sec)
Signal recognition particle (SRP)
Twin-arginine translocation (Tat)
Green fluorescent protein (GFP)
Beta-lactamase (Bla)
Beta-lactamase inhibitor protein (BLIP)
Maltose binding protein (MBP)
Trimethyl amine N-oxide reductase (TorA)
Signal peptide/signal sequence (ss)
Protein of interest (POI)
Ampicillin (Amp)
Wildtype (wt)
Alanine (Ala) (A)
Arginine (Arg) (R)
Asparagine (Asn) (N)
Aspartic acid (Asp) (D)
Cysteine (Cys) (C)
Glutamic acid (Glu) (E)
Glutamine (Gln) (Q)
Glycine (Gly) (G)
Histidine (His) (H)
Isoleucine (Ile) (I)
Leucine (Leu) (L)
Lysine (Lys) (K)
Methionine (Met) (M)
Phenylalanine (Phe) (F)
Proline (Pro) (P)
Serine (Ser) (S)
Threonine (Thr) (T)
Tryptophan (Trp) (W)
Tyrosine (Tyr) (Y)
Valine (Val) (V)
Interfering RNA (RNAi)
Colony forming units (CFUs)
Minimum inhibitory concentration (MIC)
Minimum bacteriocidal concentration (MBC)
Chloramphenicol (Cm)
Kanamycin (Kan)
Luria-Bertani (LB)
Isopropyl-beta-D-thiogalactopyranoside (IPTG)

Fluorescence activated cell sorting (FACS)
Dithiothreitol (DTT)
Consolidated bioprocessing (CBP)
Type II secretion system (T2SS)
Type III secretion system (T3SS)
Clustered regularly interspaced short palindromic repeats (CRISPR)
Cholera toxin (CT)
Heat-labile enterotoxin (LT)
Enterotoxigenic *E. coli* (ETEC)
Plant cell wall degrading enzymes (PCWDEs)

CHAPTER 1

EXTRACELLULAR SECRETION OF RECOMBINANT PROTEINS IN *ESCHERICHIA COLI*

1.0 Introduction

The gram-negative bacterium *Escherichia coli* is widely used for both laboratory- and industrial-scale biosynthesis of proteins. For decades, the strategy for preparative protein production in *E. coli* has been by overexpression of plasmid-encoded proteins for synthesis in the cytoplasmic compartment, with the optional step of secretion into the surrounding periplasm. Due to a lack of a dedicated extracellular secretion system, most proteins produced in *E. coli* are retained within the cell, either within the cytoplasmic or periplasmic compartment. Preparative protein expression from the intracellular space creates strong bottlenecks that reduce yield and increase costs associated with their production. For example, the intracellular environment of the bacterial cell is highly crowded with macromolecules, often reaching a concentration of 300-400 mg/mL (Ellis and Hartl, 1996). In this molecularly crowded environment, overexpressed proteins can easily misfold, aggregate, and form inclusion bodies, which are difficult to recover (Georgiou, 1988; Baneyx and Mujacic, 2004). The intracellular environment also contains at least seventy four proteases (Gottesman, 1996), some of which are known affect the stability of recombinant proteins of interest (POI) (Baneyx and Mujacic, 2004). Finally, intracellular protein synthesis invariably requires costly downstream purification procedures to isolate the POI from contaminating cellular debris, such as housekeeping proteins, and also lipopolysaccharides, which can be highly immunogenic.

An alternative strategy for the biosynthesis of POI is to target them for export out of cells. This would overcome many of the problems associated with intracellular protein synthesis, because, unlike the intracellular space, the extracellular space is a highly controlled environment with a very low concentration of contaminating cellular components. This is because laboratory *E. coli* strains do not secrete many host proteins to the extracellular space, including proteases. Therefore, purification of POI from the culture medium rather than from cells promises increased yield, simpler purification, and reduced costs associated with their production. Furthermore, certain industrial applications require extracellular secretion of target proteins. For example, biosynthesis of POI with cytotoxic activity may only be achieved by extracellular production. Also, extracellular secretion is required in consolidated bioprocessing (CBP) systems. CBP systems are engineered microbes that are capable of secreting plant cell wall degrading enzymes for hydrolysis of plant biomass feedstocks, and also converting the released sugars to energy products (see subchapter 1.4 for more information on CBP). Extracellular secretion of enzymes is required in these systems because cellulose is not transported into the bacterial cell.

Despite all the advantages of extracellular secretion, intracellular protein production has been the norm because *E. coli* lacks a dedicated secretion system to deliver heterologous proteins to the extracellular environment. The structural organization of the gram-negative cell means that there are two tightly sealed lipid membranes which proteins must traverse if they are to reach the extracellular space. All protein biogenesis occurs in the cytoplasmic compartment, and this is enclosed by the cytoplasmic or inner membrane, and also the periplasmic compartment, which itself is enclosed by the outer membrane. Following synthesis in the cytoplasm, a subpopulation of proteins traverse the inner membrane by a well-understood process, either by the Sec or the cotranslational signal recognition particle (SRP) system for unfolded or semi-unfolded proteins,

or the twin-arginine (Tat) translocation system for fully folded proteins. To be targeted for secretion by either of these systems, all proteins must contain a defined N-terminal signal sequence.

Extracellular secretion of proteins is less well-understood. For proteins to reach the extracellular space, they can follow one of two general secretory mechanisms (Rêgo et al., 2010). The one-step secretory mechanism, such as type I, type III, type IV, and type VI secretion, delivers proteins from the cytoplasm directly to the extracellular environment in a single step, bypassing an intermediate translocation step in the periplasmic compartment. By contrast, two-step secretion first delivers the exoprotein to the periplasmic compartment by either the Sec or Tat machinery. Once inside the periplasm, the protein can traverse the outer membrane in a separate secretion step, such as by the type II secretion system (T2SS), chaperone-usher system (CU), or other mechanism. In two-step secretion processes, the secretory protein undergoes residence time within the periplasm. In this manner, the periplasmic intermediate of the secretory protein can undergo posttranslational mechanisms exclusive to this compartment, such as disulfide bond formation and *N*-linked glycosylation (Fisher et al., 2011).

The fact that secretory proteins undergo residence time in the periplasm makes two-step secretion pathways attractive for biological and biotechnological contexts that rely on periplasm-associated posttranslational modifications for activity. For example, many enzymes secreted by T2SS contain intramolecular disulfide bonds (Korotkov et al., 2012), and many industrial proteins produced in biological hosts require disulfide-bonds for activity, e.g. antibodies. Also, many biologics on the market today are glycosylated proteins. *E. coli* is normally not capable of glycosylating proteins, but the *N*-linked glycosylation machinery from *Campylobacter jejuni* has been functionally transferred to *E. coli* cells, enabling these cells to selectively glycosylate

proteins at defined asparagine residues (Wacker et al., 2002). This protein glycosylation system functions natively in the periplasmic compartment (Kelly et al., 2006). Thus, under these conditions, glycosylated proteins can only be targeted for extracellular secretion by two-step secretion processes.

1.1 The extracellular proteome of E. coli

In the vast literature of *E. coli* biology, there have been, until recently, very few reports on extracellular secretion in *E. coli* K-12. Unlike pathogenic strains, such as enterotoxigenic *E. coli* (ETEC), which rely on secretion systems to disperse virulence factors during host colonization (Tauschek et al., 2002), non-pathogenic strains of *E. coli* secrete very few proteins to the extracellular milleau. However, in recent years, a growing body of evidence challenges the assumption that *E. coli* is a poor secretor of proteins. For example, proteomic studies in laboratory *E. coli* strains showed a number of proteins in the supernatant fraction that appeared to be secreted by mechanisms that are independent of cell lysis (Rinas and Hoffmann, 2004; Nandakumar et al., 2006; Xia et al., 2008). Most of the proteins were outer membrane and periplasmic proteins, but also included cytoplasmic proteins (Nandakumar et al., 2006; Xia et al., 2008). The proteins identified in these studies are of particular interest because they represent a potential mechanism for secreting heterologous proteins to the extracellular space. Indeed, two outer membrane proteins OmpA, OmpF, and the osmotically inducible protein Y, OsmY, were exploited as carriers for secreting a diverse range of proteins, including human and animal proteins, and also cellulases (Ni and Chen, 2009; Bokinsky et al., 2011; Kotzsch et al., 2011). The mechanisms by which these proteins are secreted into the culture medium is unknown. It is possible that outer membrane proteins OmpA and OmpF are shed or released into the

extracellular environment as soluble proteins or associated with vesicles. Due to structural and biochemical differences, the mechanism for OsmY secretion is likely different (Kotzsch et al., 2011).

One of the earlier studies to demonstrate secretion of heterologous proteins as 'passengers' on extracellular carrier proteins employed a protein called YebF. The YebF protein is a small protein (~13 kD) native to *E. coli* with unknown function, though it has structural similarity with the colicin M immunity protein (Cmi), which protects cells from the bacteriolicin colicin M (Cma) protein (Usón et al., 2012). Interestingly, there is evidence that YebF accumulates at high levels in the supernatant when overexpressed (Zhang et al., 2006). For this reason, the YebF protein can potentially be exploited to carry with it heterologous proteins from the cytoplasm all the way to the culture medium. In fact, a powerful feature of the YebF system is that it tolerates C-terminal fusions with proteins of diverse structural and functional properties. Functional enzymes ranging from 15 to 48 kD have been secreted by YebF (Zhang et al., 2006), and more recently *N*-linked glycoprotein domains (Fisher et al., 2011). Therefore, a unique feature that the YebF protein has in common with OsmY, OmpA, and OmpF is that it can serve as a 'carrier' domain for delivering POI to the culture medium without extensive genetic manipulation to the *E. coli* host. These secretion systems can also be called single domain secretion systems because they can be genetically fused as single domains to heterologous proteins for delivery to the extracellular medium.

The model for YebF secretion is a two-step process. YebF is secreted into the periplasm by the Sec system, and then it is targeted for translocation across the outer membrane by passing through the outer membrane porin OmpF, and possibly also OmpC (Fig. 1.1) (Prehna et al., 2012). The protein OmpC has structural and functional similarity to OmpF, and it appears to be

involved, though not required, in YebF outer membrane translocation. It seems likely that YebF can be transported through an OmpC/OmpF complex since OmpC has previously been demonstrated to form stable heterotrimers with OmpF (Kumar et al., 2010). Finally, a third component, OmpX, is also required for YebF secretion, possibly by targeting YebF to the OmpF/C complex (Fig. 1.1) (Prehna et al., 2012). While outer membrane porins have long been known to import bacteriolicins and other molecules (Kleanthous, 2010), this study provides compelling evidence that endogenous proteins like YebF can be exported out of cells by outer membrane porins.

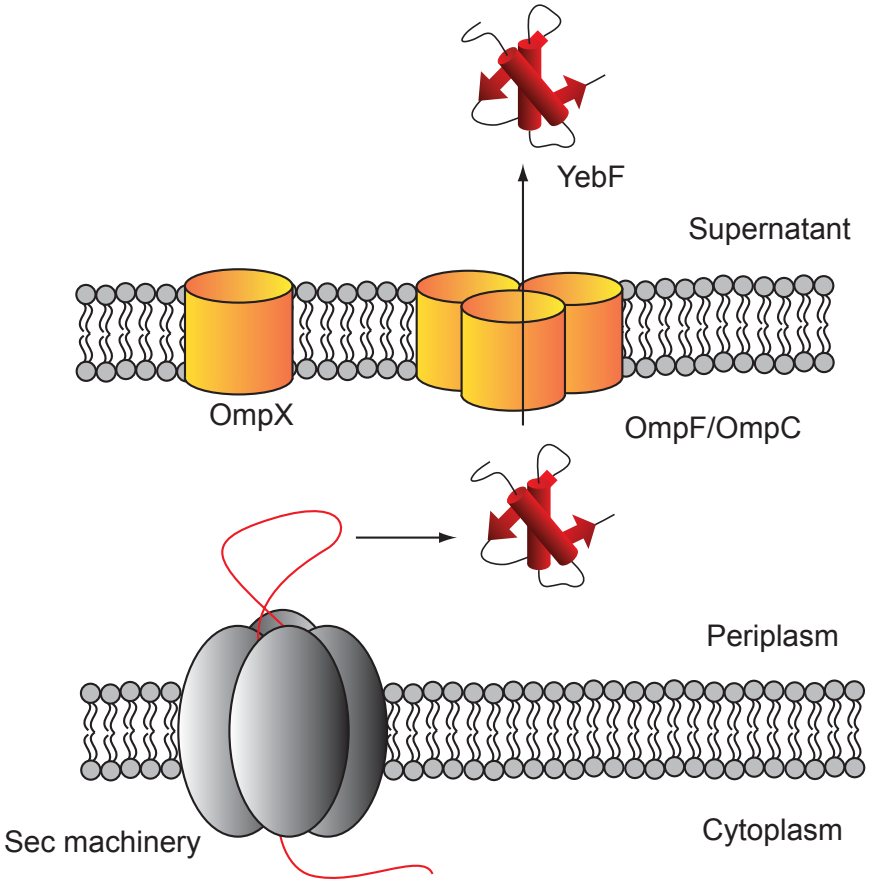

Figure 1.1 Model for YebF translocation across the outer membrane. YebF is translocated through outer membrane porins OmpF, which forms stable homotrimers, or heterotrimers with OmpC. Outer membrane protein OmpX is required for translocation, possibly by directing YebF to the OmpF/OmpC complex. Adapted from (Prehna et al., 2012).

1.2 Type II secretion

The studies mentioned thus far indicate that endogenous proteins like YebF, and also OsmY, OmpA, and OmpF can be harnessed to serve as 'carrier' domains to secrete heterologous proteins as 'passenger' domains. An alternative strategy for heterologous protein secretion in *E. coli* expression hosts is to transfer exogenous secretion machinery from other bacteria. Multiple studies have demonstrated that exogenous T2SS systems can be expressed in *E. coli* to secrete canonical substrates that pass through these systems in their native hosts. For example, T2SS machinery have been functionally transferred from *Klebsiella oxytoca* (Michaelis et al., 1985) and *Dickeya dadantii* (He et al., 1991), enabling *E. coli* to secrete pullanase and pectinase, respectively. Typically, when expressed in the absence of T2SS machinery, exoproteins localize to the periplasm because *E. coli* lacks the machinery to deliver these proteins across the outer membrane (Pugsley, 1993). However, it should be noted that *E. coli* K-12 has an endogenous and functional T2SS encoded by the cryptically-expressed *gsp* operon (Francetic and Pugsley, 1996). These genes are named after the general secretory pathway (GSP) because it is considered to be the main terminal branch of the GSP (Francetic and Pugsley, 1996). However, there are two bottlenecks that limit the exploitation of this *E. coli* system to be harnessed for heterologous protein secretion. First, while some T2SS can secrete a dozen or more proteins (Kazemi-Pour et al., 2004), the only known substrate to be secreted by the *gsp* T2SS is a chitinase encoded by *chiA*. Second, expression of *gsp* T2SS is negatively controlled by the universal repressor protein H-NS. Thus, the *gsp* operon is not expressed under laboratory conditions. Only when H-NS protein is removed and the *gsp* operon is overexpressed can ChiA be secreted in detectable amounts (Francetic et al., 2000a).

Like the YebF secretion system, T2SS is a two-step secretion process; substrates targeted for T2SS enter the periplasm by Sec or Tat machinery, and then traverse the outer membrane by the T2SS assembly. Proteins translocated through the outer membrane are fully folded and often contain intramolecular disulfide bonds (Korotkov et al., 2012). The T2SS is encoded by 12-15 genes, usually arranged in an operon (Korotkov et al., 2012). These genes are conserved in many gram-negative organisms, and they encode a macromolecular complex that spans the inner membrane, periplasm, and outer membrane. The inner membrane platform contains an ATPase that provides energy for T2SS-mediated translocation. In this working model of T2SS, a filamentous structure directly interacts with exoproteins and polymerizes in a pilus-like assembly to 'push' exoproteins through an outer membrane porin encoded by *E. coli gspD* (Korotkov et al., 2012).

A long-standing question in T2SS biology is how exoproteins are recognized by the T2SS assembly. No linear amino acid signal sequence has been identified, and one possibility is that the signal sequence is encoded within the three-dimensional confirmation of the exoprotein (Korotkov et al., 2012). Protein domain deletion experiments identified a conserved PDZ domain in the *D. dadantii* OutC protein (*E. coli* GspC) that is required for secretion activity and also appears to interact directly with exoproteins (Bouley et al., 2001), indicating that this motif may be at least partly responsible for determining substrate specificity. Furthermore, OutC, and also OutD (*E. coli* GspD), are the only proteins of the *out* operon that cannot be complemented by homologs from the closely related organism *Pectobacterium carotovorum*, suggesting these two proteins are the 'gatekeepers' of substrate specificity among different species (Lindeberg et al., 1996). These unresolved aspects may limit the exploitation of T2SS for heterologous protein secretion.

Many plant and animal pathogens rely on T2SS for host invasion. Some toxins of human pathogens are secreted by T2SS, like the cholera toxin (CT) of *Vibro cholerea* (Sandkvist et al., 1997) and the heat-labile enterotoxin (LT) of enterotoxigenic *E. coli* (ETEC) (Tauschek et al., 2002). The T2SS also mediates the secretion of enzymes with highly degradative effects on plant and animal tissues, such as phospholipases, proteases, and lipolytic enzymes secreted by *Psuedomonas aeurginosa*, the causative agent of cystic fibrosis (Jyot et al., 2011). This is particularly true for plant pathogens which use T2SS to secrete a broad range of plant cell wall degrading enzymes (PCWDEs) (DeBoy et al., 2008),(Lindeberg, 1992), including pectinases and cellulases (Table 1.2). This feature makes T2SS a potentially powerful method for secreting cellulase enzymes and other PCWDEs in engineered CBP microbes. This speculation is further supported by the recent observation that T2SS is required for *Cellvibrio japonicus* to grow on model pretreated plant biomass feedstocks, like corn stover and switchgrass (Gardner and Keating, 2010). Furthermore, T2SS from the phytopathogen *D. dadantii*, which secretes at least 13 PCWDEs via this system (Kazemi-Pour et al., 2004), has been functionally transferred to *E. coli* cells (He et al., 1991). These two studies may be useful for future studies in engineering CBP microbes with T2SS machinery.

Name of bacterium	Known effectors
Pectobacterium carotovorum	Pectate lyases Cellulase/endoglucanase Polygalacturonase
Dickeya dadantii	Pectate lyases Cellulase/endoglucanase Polygalacturonase Pectin acetylesterase Pectin methylesterase Rhamnogalacturonate lyase FaeD esterase
Ralstonia solancearum	Polygalacturonase Cellulase/endoglucanase Pectin methyl esterase Cellobiohydrolase
Xanthomonas campestris	Polygalacturonase Cellulase/endoglucanase Proteases Alpha-amylase
Xanthomonas orzae	Xylanase Celulase/endoglucanase Putative cysteine protease Putative cellobiohydrolase Lipase/esterase
Cellvibrio japonicus	Cellulases/endoglucanase Beta-glucosidase Cellobiohydrolase

Table 1.2 Type II secretion system-secreted protein from phytopathogenic bacteria. Plant pathogens rely on the T2SS to secrete PCWDEs. Many PCWDEs are virulence factors that help pathogens colonize their host, causing diseases in plants such as bacterial soft rot by *Dickeya* species and bacterial wilt by *R. solanacearum*. Adapted from (Sonti, 2005).

1.3 Type III secretion

Like the T2SS, the type III secretion system (T3SS) is conserved in many gram-negative bacteria, and it is used by plant and animal pathogens to secrete virulence factors. But unlike the T2SS, T3SS is a one-step secretion mechanism that is capable of delivering proteins from the bacterial cytoplasm directly to the extracellular medium, or into the cytosol of plant and animal

cells, thereby bypassing intermediate residence of the exoprotein within the periplasm. (Cornelis, 2006). The injecting of effector proteins into eukaryotic cells is mediated by an injectosome, which is a highly complex and highly regulated macromolecular nanomachine composed of about 25 proteins (Hodgkinson et al., 2009). These proteins assemble to form an ultrastructure protruding from the cellular envelope that resembles a molecular 'syringe,' and these proteins are genetically, morphologically, and functionally related to flagellar basal bodies (Blocker et al., 2003).

The T3SS has advantages over other secretory mechanisms in certain contexts within biotechnology. First, the unique ability of T3SS to inject effector proteins in eukaryotic cells can been exploited as a novel approach for targeting therapeutics into host cells, and this has already be done in *Salmonella* and *Yersinia* species (Rüssmann et al., 1998; Boyd et al., 2000; Konjufca et al., 2006). In principle, the same approach could be used to treat diseases in humans. For instance, the non-pathogenic *E. coli* strain Nissle 1917 has been widely used as a probiotic agent to treat inflammation and infection of the intestine (Lodinová-Žádníková et al., 1998), and has recently been used to deliver therapeutic molecules by surface display (Westendorf et al., 2005), and by the Hly type I secretory apparatus (Rao et al., 2005). Thus, engineering *E. coli* Nissle 1917 with T3SS may one day be a powerful strategy for drug delivery. Although *E. coli* K-12 and Nissle are not known to encode a T3SS, a cloned T3SS called Hrp from *Dickeya dadantii* has been shown to function in *E. coli* cells (Ham et al., 1998).

A second advantage of T3SS over T2SS is that the N-terminal signal tag, chaperone, and chaperone binding domain for proteins secreted by the T3SS have been identified (Galán and Collmer, 1999), and this information has been used to secrete recombinant proteins via this pathway (Widmaier et al., 2009). Despite these advantages, a limitation to the T3SS is that the

exoproteins are delivered from the cytoplasm to the extracellular space in one step, thus proteins that depend on periplasm-associated posttranslational modifications, like glycosylation and disulfide bond formation, cannot be exported by the T3SS.

1.4 Consolidated bioprocessing

Cellulose is the most abundant carbon compound on earth and it represents an untapped reservoir for bioenergy. Currently, biochemical processing of plant biomass requires the addition of costly cellulolytic enzymes to overcome the recalcitrant nature of cellulose (Fig. 1.4A). One strategy for lowering costs is to engineer microbe-based production systems for the synthesis and extracellular secretion of cellulases. These microbes can in principle be engineered for consolidated bioprocessing (CBP), whereby the secretion of the plant cell wall degrading enzymes and fermentation of released sugars to target energy products like ethanol occurs in a single host organism (Fig. 1.4A). *Sacchromyces cerevisiae* is well-known ethanogen, but gram-negative bacteria, such as *Zymomonas mobilis*, are also capable of commerical-scale ethanol production (Rogers et al., 2007). *E. coli* has also been engineered for efficient ethanol production (Jarboe et al., 2007), and other bioenergy products like 1,2-propanediol (Altaras and Cameron, 1999), butanol, pinene, and fatty acid ethyl ester (Bokinsky et al., 2011). A major challenge in the development of gram-negative bacteria for CBP is the lack of a robust secretion system for targeting PCWDEs capable of degrading plant biomass feedstocks. Ideally, this secretion system would be able to recognize and secrete heterologous enzymes engineered for optimal performance on industrially relevant plant biomass feedstocks. A closely related organism *Cellvibrio japonicus* is well known for degrading plants (DeBoy et al., 2008), and it can grow on industrial feedstocks like pretreated corn stover and switch grass (Gardner and Keating, 2010).

Thus, one strategy is to metabolically engineer organisms already capable of plant biomass degradation, like *C. japonicus,* with metabolic pathways for converting the released sugars to target chemical products. An alternative strategy is to engineer dedicated secretion platforms for secretion of cellulases in *E. coli*, so that these cells have the ability to use biomass feedstocks for production of biofuels. This approach was taken in a recent study, which relied on the OsmY secretion system to secrete cellulases and convert biomass feedstocks to three model biofuels (Fig. 1.4B) (Bokinsky et al., 2011).

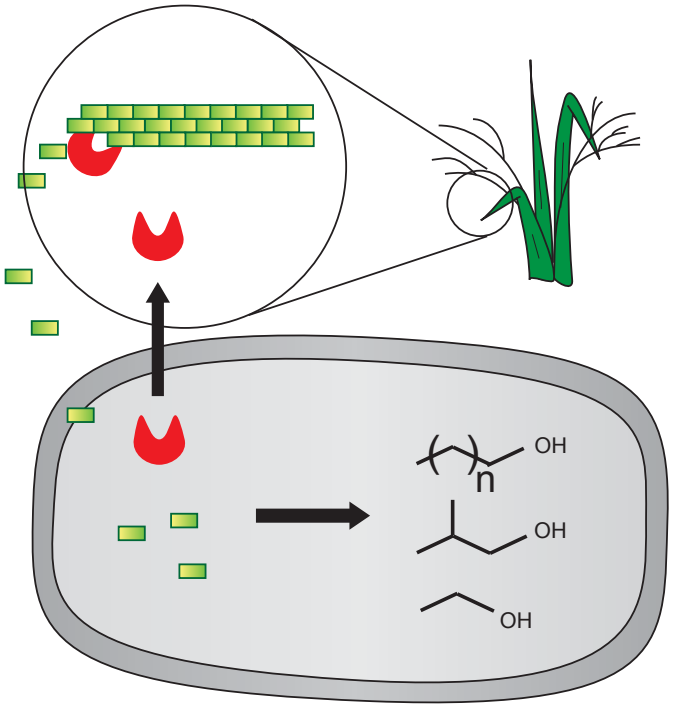

Figure 1.4 Consolidated bioprocessing. In consolidated bioprocessing systems, enzyme generation, biomass hydrolysis, and biofuel production are performed in a single step. Microbes capable of consolidated bioprocessing secrete cellulases and convert released sugars to target bioenergy products. Cellulases (circles) synthesized by the CBP microbe release monsaccharides (rectangles) from plant biomass, which can be fermented by the CBP microbe to to target energy products, such as ethanol, isobutanol, or fatty alcohols. Adapted from (Bokinsky et al., 2011).

CHAPTER 2

PRODUCTION OF EXTRACELLULAR *N*-LINKED GLYCOPROTEINS IN *ESCHERICHIA COLI* [1]

2.0 Introduction

Given the recent success in *exporting* recombinant proteins via the YebF pathway (Zhang et al., 2006), we investigated the ability of this system to secrete therapeutically and industrially relevant biomolecules. Many therapeutic biologics on the market today are glycosylated proteins. Asparagine-linked (*N*-linked) protein glycosylation is essential and conserved in eukaryotic organisms. It is the most prevalent of all posttranslational protein modifications, affecting nearly 70% of the eukaryotic proteome (Apweiler et al., 1999). The attachment of N-glycans to eukaryotic secretory and membrane proteins can influence their folding and stability, oligomerization, resistance to proteolysis, sorting, and transport (Helenius and Aebi, 2001; 2004). *N*-linked glycosylation occurs in the endoplasmic reticulum (ER) and involves the assembly of glycans on a lipid carrier in the ER membrane followed by transfer to specific asparagine residues of target polypeptides. Initially, it was believed that *N*-linked glycosylation was unique to eukaryotes. However, *N*-glycoproteins have now been described for all domains of life, including archaea and more recently bacteria, of which the best-characterized example is the human gastroenteric pathogen *Campylobacter jejuni* (Szymanski and Wren, 2005). In *C. jejuni*,

---

[1]Adapted with permission from:

Fisher A. C., **C. H. Haitjema**, C. Guarino, E. Celik, C. E. Endicott, C. A. Reading, J. H. Merritt, A. C. Ptak, S. Zhang, and M. P. DeLisa. 2011. Production of secretory and extracellular *N*-linked glycoproteins in *Escherichia coli*. Appl Environ Microbiol, 2010th ed. **77**:871–881.

the genes for this pathway comprise a 17-kb locus named *pgl* for protein glycosylation (Szymanski et al., 1999). To date, more than 40 periplasmic and membrane glycoproteins have been identified in *C. jejuni* (Young et al., 2002; Kowarik et al., 2006b), and most of these bind to the N-acetyl galactosamine (GalNAc)-specific lectin soybean agglutinin (SBA) (39). Mass spectrometry and nuclear magnetic resonance (NMR) studies revealed that the *N*-linked glycan is GlcGalNAc5 Bac, where Bac is bacillosamine (2,4-diacetamido- 2,4,6-trideoxyglucose) (Young et al., 2002). This branched heptasaccharide is synthesized by sequential addition of nucleotide-activated sugars on the lipid carrier undecaprenyl pyrophosphate on the cytoplasmic face of the inner membrane (Feldman et al., 2005). Once assembled, the lipid-linked heptasaccharide is flipped across the membrane by the putative ATP-binding cassette (ABC) transporter PglK (Alaimo et al., 2006; Kelly et al., 2006). Transfer of the heptasaccharide to periplasmic substrate proteins is catalyzed by an oligosaccharyltransferase (OST) called PglB, a single integral membrane protein with significant sequence similarity to the catalytic subunit of the eukaryotic OST STT3 (Young et al., 2002). PglB attaches the heptasaccharide to asparagine in the motif D/E-X1-N-X2- S/T (where X1 and X2 are any residues except proline), a sequon similar to that of eukaryotes (Kowarik et al., 2006b).

Recently, Wacker and coworkers transferred the entire *C. jejuni pgl* locus into *Escherichia coli*, conferring upon these cells the ability to *N*-glycosylate proteins (Wacker et al., 2002). Native *C. jejuni* glycoproteins such as Peb3 and AcrA, which are localized to the periplasm by the Sec pathway, can be *N*-glycosylated in glycosylation-competent *E. coli* (Wacker et al., 2002). AcrA can also become *N*-glycosylated when transported via the twin-arginine translocation (Tat) pathway, which is well known for its ability to export folded proteins across the inner membrane (Kowarik et al., 2006a). In addition to periplasmic proteins, some native *C.*

jejuni N-glycoproteins are predicted to be integral membrane proteins based on bioinformatic analysis (Kowarik et al., 2006b). Collectively, these earlier studies suggest that the *C. jejuni N*-linked glycosylation machinery is compatible with diverse secretory mechanisms and can tolerate an array of structures ranging from unfolded polypeptides to completely folded (albeit highly flexible and solvent-exposed) protein domains. However, the influence of membrane translocation and folding of acceptor proteins on bacterial *N*-linked glycosylation has not been thoroughly addressed. It is also not known whether proteins destined for locations beyond the periplasm, such as the outer membrane or the extracellular medium, are compatible with *N*-linked glycosylation.

Therefore, the goal of this study was to investigate the extent to which different secretory and extracellular protein substrates could be *N*-glycosylated in *E. coli* cells carrying the pgl locus. To address this issue, we developed a genetically encoded N-glycan acceptor peptide tag (GT) that can be appended terminally or inserted at internal locations of recombinant proteins. Numerous recombinant proteins modified with the GT were reliably glycosylated in *E. coli* strains expressing the *pgl* genes. When the GT was used in combination with proteins targeted to the periplasm by different export pathways (e.g., Sec, signal recognition particle [SRP], or Tat), we observed a clear difference in the glycosylation patterns on these proteins depending on their mode of inner membrane translocation. In all cases tested, N-glycan attachment via the GT did not have any measurable effect on the protein's activity. Finally, we show that proteins targeted to different locations, including the periplasm, the outer membrane, membrane vesicles, and the extracellular medium, were amenable to *N*-linked glycosylation.

2.1 Extracellular production of extracellular N-linked glycoproteins in E. coli

Since vesicle-mediated secretion was compatible with *N*-linked glycosylation (Fisher et al., 2011), we next determined whether proteins glycosylated in the periplasm could be directed for secretion across the outer membrane and into the culture medium. We chose the *E. coli* YebF protein as a carrier to target proteins for extracellular secretion. YebF is a small protein (10.8 kDa in the mature form) that is transported into the bacterial periplasm by the Sec translocase and then delivered across the outer membrane by an unknown mechanism (Zhang et al., 2006). YebF can carry C-terminal fusion partners into the culture medium at high levels (Zhang et al., 2006). To determine if *N*-glycosylated derivatives of YebF could be secreted into the culture supernatant, we created a YebF-GT chimera in plasmid pTrc99A. Extracellular YebF-GT was secreted from both pgl and pglmut cells and reacted readily with anti-His antibodies (Fig. 2.1a). The secreted YebF-GT from pgl cells, but not pglmut cells, also reacted with hR6P (Fig. 2.1a). As expected, YebF-GT lacking its native Sec signal peptide (ΔspYebF-GT) was not detected in the supernatant fraction (Fig. 2.1a).

To determine if YebF could carry recombinant glycoproteins into the supernatant, we translationally fused ΔspMBP-GT to the C terminus of YebF. The YebF-ΔspMBP-GT fusion was secreted into the culture medium by both pgl and pglmut cells, but only the fusion protein secreted from pgl cells was observed to react with hR6P (Fig. 2.1b). For secretion controls, we tested ΔspYebF-ΔspMBP-GT, which lacks an N-terminal export signal, and spYebF-ΔspMBP-GT, which modifies ΔspMBP-GT with only the native YebF signal peptide and does not include the mature YebF domain. As expected, ΔspYebF-ΔspMBP-GT was not localized in the periplasm or the supernatant (Fig. 2.1b) owing to the absence of an N-terminal signal peptide. The spYebF-ΔspMBP-GT construct was exported into the periplasm of glycosylation-competent

cells, where it became glycosylated; however, only a faint band corresponding to the fusion was detected in the culture medium (Fig. 2.1b), indicating that the YebF protein is required for extracellular secretion of the fusion proteins tested here.

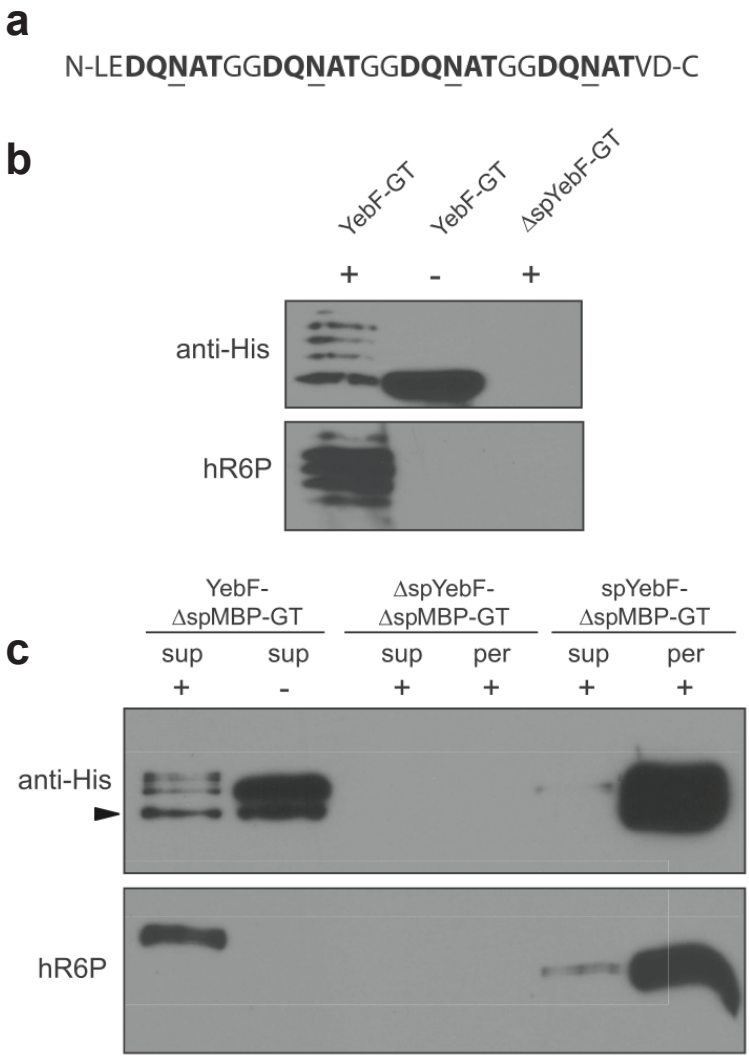

Figure 2.1 Secretion of *N*-glycoproteins in the culture supernatant. (a) The GT is comprised of four consecutive D-X1-N-X2-T sequons that are efficiently glycosylated in bacteria. (b) Western blot analysis of YebF-GT and ΔspYebF-GT in the culture supernatant of MC4100 cells. (c) Western blot analysis of ΔspMBP-GT fused to full-length YebF, ΔspYebF, and the spYebF signal peptide as indicated. The arrowhead indicates the YebF– ΔspMBP-GT degradation product. The supernatant (sup) and periplasmic (per) fractions were probed with anti-His antibodies or hR6P serum as indicated. Proteins were expressed in cells carrying pACYCpgl (+) or pACYCpglmut (-). All proteins were Ni-NTA affinity purified from the supernatant and periplasmic fractions. The same amount of protein was loaded in each lane. Blots were probed with anti-His or hR6P antibodies as indicated.

2.2 Discussion

In this study, we have demonstrated the *N*-linked glycosylation of diverse secretory and extracellular protein substrates expressed in *E. coli* cells carrying the *C. jejuni* pgl locus, thereby expanding the *N*-linked glycome of recombinant *E. coli*. This was achieved by the development of a genetically encoded glycosylation tag that promoted reliable *N*-linked glycosylation when introduced at the termini or at internal locations of recombinant proteins. Using this tag, proteins targeted to different locations, including the periplasm (via the Sec, SRP, and Tat pathways), the outer membrane, membrane vesicles, and the extracellular medium, were efficiently glycosylated. To our knowledge, this is the first report that the *C. jejuni N*-glycosylation machinery is compatible with cotranslational (i.e., SRP), outer membrane, and extracellular protein targeting mechanisms in *E. coli*.

These results also illuminate key mechanistic similarities and differences between prokaryotic and eukaryotic *N*-linked glycosylation. In eukaryotes, N-glycan addition occurs on polypeptides as they emerge through the translocation pore, a process called cotranslocational glycosylation. In this scenario, acceptor proteins are translocated into the ER in an extended conformation, and the OST is thought to recognize polypeptides in an unfolded state (Whitley et al., 1996; Chen and Helenius, 2000). In the bacterial system, translocation and glycosylation are apparently uncoupled, and glycosylation sites are located in flexible parts of folded proteins (Kowarik et al., 2006a). Indeed, we found that protein substrates exported in a folded conformation by the Tat pathway were *N*-glycosylated in the periplasm, consistent with recent findings of Kowarik and coworkers (Kowarik et al., 2006a), who also showed that completely folded proteins can be modified by PglB in vivo. We also found that the bacterial OST can

accommodate unfolded proteins, as well as folded proteins, as substrates. For example, outer membrane proteins, which are exported by the Sec pathway and maintained in an unfolded conformation until insertion in the outer membrane (Bos et al., 2007), were efficiently glycosylated by PglB. Presumably, glycosylation occurred in the periplasm prior to outer membrane integration, when folding was still incomplete. Likewise, the vesicle-mediated and extracellular secretion mechanisms investigated here, although poorly understood, are likely to involve unfolded periplasmic intermediates that are amenable to glycosylation by PglB. Taken together, our results reveal PglB to be a very versatile enzyme based on its ability to glycosylate both folded and unfolded substrates. While the glycosylation of Tat substrates here and elsewhere (Kowarik et al., 2006a) clearly indicates that PglB attaches N-glycans posttranslationally, our results do not exclude the possibility that PglB glycosylates some substrates in a cotranslational manner. In support of this possibility, we observed that proteins targeted to the SRP pathway were competent for glycosylation, and we are now determining whether these proteins undergo cotranslational glycosylation. Such a coupling between translocation and glycosylation would have important consequences for bacterial glycoengineering because potential acceptor sites would not be confined to locally flexible structures but instead could be located in structured domains that become glycosylated prior to folding.

An important outcome of our studies is the demonstration that a simple glycosylation tag is sufficient to promote covalent attachment of multiple glycans to a single protein carrier. We anticipate that this tag will be useful in a number of important areas related to glycobiology and glycomedicine. For example, the GT could be used to express and isolate carrier proteins that are hyper-glycosylated with a multiplicity of homogeneous glycans. Such hyper-glycosylated proteins could then be used as features on carbohydrate microarrays (so-called glycoarrays),

which are currently very challenging to fabricate due in large part to difficulties associated with the synthesis and isolation of sufficient quantities of naturally occurring oligosaccharides (Laurent et al., 2008). Glycan diversity could be "genetically encoded" by expression of different glycosyltransferases to control the specific glycoform, and covalent transfer onto target proteins could be achieved using PglB, which is relatively promiscuous in its choice of both oligosaccharide (Feldman et al., 2005) and protein (Fig. 3) substrates. Another potential use of the GT is in the development of therapeutic glycoprotein conjugates. For example, bacterial polysaccharides conjugated to proteins have proven effective as vaccines, as evidenced by the Haemophilus influenzae type b conjugate vaccine (Verez-Bencomo et al., 2004). By expanding the spectrum of recombinant protein carriers that can be N glycosylated via fusion to the GT, more-complex conjugate vaccine candidates may be developed. For example, multiple doses of H. influenzae type b capsular polysaccharide conjugate vaccines are required to induce protective antibody responses in infants. When conjugated to the meningococcal outer membrane protein (OMPC), protective antibody responses are induced after a single dose (Pérez-Melgosa et al., 2001). A recombinant glycosylation tag could allow for *N*-linked glycosylation of optimized protein carriers in *E. coli*. Even MBP could serve as a useful carrier for glycoconjugate vaccines since (i) it can be engineered to carry far more glycans than most naturally occurring glycoproteins (Fig. 1b) (Ben-Dor et al., 2004) and (ii) it has demonstrated immunostimulatory properties (Fernandez et al., 2007). A final application for the GT is in the development of genetic tools for analyzing glycosylation phenotypes, which at present are severely lacking. A recombinant peptide tag allows glycosylation of proteins that may be amenable to extracellular display (Smith, 1985; Francisco and Georgiou, 1994) and/or two-hybrid systems (Joung et al., 2000) for the isolation of specific N-glycans. Further, a GT could be fused to a permissive site in

a binding protein or enzyme such that glycosylation modulates the activity of the enzyme. Such a scheme would allow for genetic selection or screening dependent on functional *N*-linked glycosylation. It is our hope that the glycosylation tag strategy presented here allows for a higher level of sophistication and throughput in the emerging field of glycoengineering and opens the door to bacterial synthesis of a wide array of recombinant glycoprotein conjugates.

2.3 Materials and methods

2.3.1 Bacterial strains and growth conditions

E. coli strain DH5α was used for cloning of plasmids. For YebF-mediated secretion studies, strain MC4100 was used because this strain exhibits the least leakage of maltose-binding protein (MBP) of 15 common *E. coli* strains tested (Zhang et al., 2006), and a derivative strain showed no leakage of MBP during its early growth stage (Francetic et al., 2000b). Overnight *E. coli* cultures were diluted in fresh Luria-Bertani broth (LB) supplemented with antibiotics and 0.2% glucose and grown at 30°C or 37°C. At mid-log phase (optical density at 600 nm [OD600], ~0.5), the medium was changed to glucose-free LB containing antibiotics and protein expression was induced with either 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) for pTrc99A-based expression vectors or 0.2% arabinose for pBAD-based expression vectors. Induction was at 25°C or 30°C for up to 24 h. Antibiotics were used at the following concentrations: 100 μg/ml ampicillin (Amp), 25 μg/ml chloramphenicol (Cm), and 50 μg/ml Kan.

2.3.2 Plasmid construction

All YebF plasmids were derivatives of pTrc99A. For these, *E. coli* yebF was PCR amplified and cloned between SacI and XbaI in pTrc99A. Controls were similarly constructed by

cloning either the YebF N-terminal signal peptide (spYebF) or the mature domain of YebF lacking the N-terminal signal peptide (ΔspYebF) between SacI and XbaI in pTrc99A. The GT or MBP-GT fusion was then added to each YebF vector between the XbaI and SalI sites. Lastly, a 6×-His affinity tag was inserted between the SalI and HindIII sites. The sequences of all plasmids constructed in this study were confirmed by DNA sequencing.

2.3.3 Subcellular fractionation and protein purification

To isolate intracellular glycoproteins, equal numbers of cells were pelleted by centrifugation at 5,000 × g for 15 min at 4°C, resuspended in 1 ml lysis buffer supplemented with 1% (vol/vol) Triton X-100 and 1 mg/ml lysozyme, and incubated on ice for 30 min. Cells were then sonicated 4 times for 30 s with a 1-min rest between each sonication. The sonicated cells were spun down at 10,000 × g at 4°C for 20 min, and the supernatants were collected. For preparation of periplasmic and culture supernatant fractions, equal numbers of cells were harvested by centrifugation at 5,000 × g for 15 min at 4°C. The supernatant fraction was subjected to 0.2-μm filtration and then concentrated with an Amicon Ultra centrifugal filter from Millipore. The cell pellet was washed and then subjected to subcellular fractionation into periplasmic and cytoplasmic fractions using the ice-cold osmotic-shock procedure as described elsewhere (DeLisa et al., 2003; Kim et al., 2005). All 6×-His-tagged proteins were purified using Ni-nitrilotriacetic acid (NTA) spin columns (Qiagen) according to the manufacturer's instructions (Ni-NTA spin kit; Qiagen) under native conditions. Fc domains were purified using NAb protein A/G spin columns according to the manufacturer's instructions (Thermo Scientific).

2.3.4 Protein analysis

Proteins were separated with SDS-polyacrylamide gels (Bio-Rad), and Western blotting was performed as described previously (DeLisa et al., 2003). Briefly, proteins were transferred onto polyvinylidene fluoride (PVDF) membranes, and membranes were probed with one of the following: anti-MBP antibodies conjugated with horseradish peroxidase (HRP) (New England Biolabs), anti-6×-His antibodies conjugated to HRP, and hR6P antiserum, which is specific for the C. jejuni heptasaccharide (kindly provided by Markus Aebi). In the case of hR6P antiserum, anti-rabbit IgG-HRP (Promega) was used as the secondary antibody.

CHAPTER 3

A UNIVERSAL APPROACH FOR ENGINEERING EXTRACELLULAR SECRETION PATHWAYS

3.0 Introduction

There is unprecedented interest in converting plant biomass into fermentable sugars for the production of bioenergy products. Currently, the biochemical processing of plant biomass requires the addition of costly cellulolytic enzymes due to the complex and crystalline nature of lignocellulose, the major component of plant biomass (Wilson, 2009). An alternative strategy is consolidated bioprocessing (CBP) whereby by a single organism secretes plant cell wall degrading enzymes and simultaneously ferments the released sugars into fuel products. A CBP platform could dramatically reduce production costs associated with cellulosic biofuels (Lynd et al., 2008). *E. coli* remains the most robust host for recombinant protein productions and can be engineered for efficient production of ethanol (Jarboe et al., 2007), but it lacks a dedicated extracellular secretion system. Extracellular secretion is a requirement for degradation of plant cell walls, because cellulose is intractable and not transported into the microbial cell. Many phytopathogenic bacteria rely on large macromolecular complexes called type II secretion systems T2SS to secrete plant cell wall degrading enzymes (Sonti, 2005). However, there are two major limitations that could prevent T2SS for use in CBP platform. First, the secretion signal for targeting heterologous proteins to this pathway has not been identified and so T2SS can only recognize cognate exoproteins (Korotkov et al., 2012). Second, there appears to be a stringent species specificity associated with T2SS that prevents these systems from secreting the proteins

of even closely related organisms (Lindeberg et al., 1996). An alternative to T2SS is a single-domain secretion system called YebF that can efficiently carry recombinant proteins into the extracellular environment (Zhang et al., 2006; Fisher et al., 2011). It was recently discovered that YebF and YebF-fusions can be glycosylated by engineered strains of *E. coli* (Fisher et al., 2011). This is an important observation as glycosylation affects the properties of many industrial enzymes including cellulases (Jeoh et al., 2008). In this report, we engineered *E. coli* K-12 for robust extracellular secretion of a recombinant cellulolytic cocktail via the YebF secretion pathway. The enzymes were rationally selected from phytopathogenic bacterium, *Cellvibrio japonicus*, which is well known for degrading plant biomass (DeBoy et al., 2008; Gardner and Keating, 2010). Secretion of these cellulases conferred the ability of non-cellulolytic *E. coli* to grow on carboxymethyl cellulose (CMC) as a sole carbon source. Finally, to increase cellulase secretion titers, we screened a transposon-insertion library that enhance extracellular accumulation of YebF. We did this by developing high-throughput screening technology for extracellular secretion based on the biarsenical FlAsH-EDT2 dye. The mutants isolated in this study exhibited enhanced export of YebF and YebF-cellulase fusions to the extracellular space. As a result, we were able to show that these strains were able to hydrolyze and grow on cellulose more robustly than the parental strain.

3.1 Co-secretion of a cellulolytic cocktail confers growth on cellulose.

E. coli lacks two key components necessary for a CBP platform, (i) cellulolytic activity encoded by cellulases and hemicellulases for hydrolysis of lignocellulose into monosaccharides that can imported and metabolized by the cell, and (ii) a secretion platform for delivering these enzymes to the extracellular environment. To address the first issue, we rationally selected four

enzymes that we predicted would represent the basic set of cellulolytic activity to efficiently degrade cellulosic and hemicellulosic substrates. The enzymes, encoded in *Cellvibrio japonicus*, a gram-negative bacterium known for efficient plant cell wall degradation (DeBoy et al., 2008), include Cel3A, Cel5B, Cel6A, and Cel9A, a predicted beta-glucosidase, endoglucanase, cellobiohydrolase, and processive endoglucanase, respectively. While *C. japonicus* requires T2SS for growth on cellulose (Gardner and Keating, 2010), it is not currently known if these enzymes are secreted by the T2SS. Thus, to target these proteins for extracellular secretion, we fused each enzyme to the C-terminus of YebF, a protein known for carrying a broad range of proteins to the extracellular environment (Zhang et al., 2006). The sequence of each enzyme was codon optimized for expression in *E. coli* and cloned into an expression vector. A C-terminal 6H affinity tag was included for downstream purification and immunoblot analysis. When these plasmids were expressed in wild-type BW25113 cells, bands corresponding to 6H were observed in the supernatant fraction, and in some cases the periplasmic fraction, during immunoblot analysis (Fig. 3.1a). These data indicated that these proteins were separately expressed and targeted for extracellular secretion by *E. coli*. We next determined if these proteins could be simultaneously co-secreted in a single strain by the YebF pathway. Thus, we created plasmid pCEL4 which encodes all for YebF-cellulase fusions expressed polycistronically with different C-terminal epitope tags for discrete detection by immunoblot analysis. When this plasmid was transformed into wild-type BW25113 cells, bands corresponding to each epitope tag appeared in the supernatant fraction. Bands corresponding to YebF-Cel5B and YebF-Cel6A were also observed in the periplasm, though they are not as strong as the same bands observed in the supernatant (Fig. 3.1b), indicating that these proteins are efficiently translocated across the outer membrane.

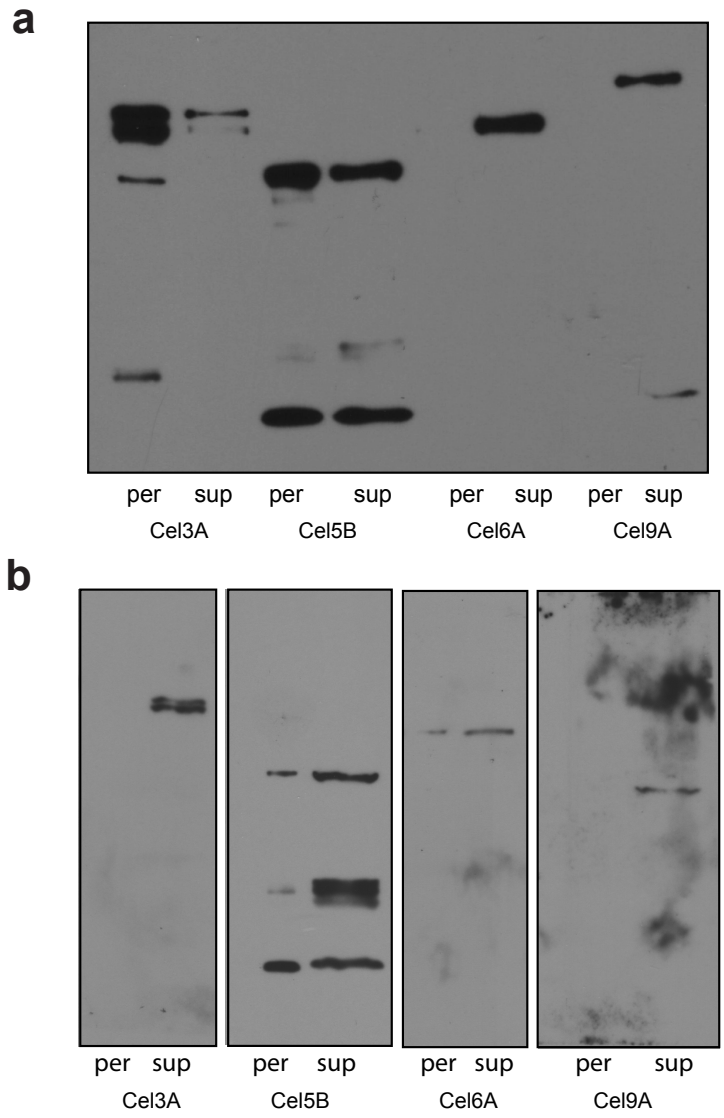

Figure 3.1 Extracellular secretion of cellulases fused to YebF. Western blot analysis of periplasmic (per) and supernatant (sup) fractions of wild-type cells expressing YebF-cellulase fusions (a) individually or (b) simultaneously. Individually expressed cellulase fusions were detected using anti-6H antibody. Simultaneously expressed YebF-cellulase fusions were detected using (from left to right) anti-cmyc, anti-FLAG, anti-6H, and anti-HA antibodies

The experiments thus far suggest that our chosen four cellulases can be simultaneously synthesized and translocated to the extracellular medium by a single *E. coli* strain. We next asked the question whether expressing these proteins conferred upon these cells the ability to consume cellulose as a sole carbon source. Wild-type strains expressing each cellulase individually, or all

four cellulases in a single strain, were inoculated in M9 minimal medium containing 0.2% glucose or 1% carboxymethyl cellulose (CMC) or 0.34% cellobiose as a sole carbon source. As expected, all cells grew in the presence of glucose. Only cells carrying YebF-Cel3A were able to grow on cellobiose, whereas cells not expressing Cel3A exhibited no measurable growth (Table 3.1.1). These data are consistent with the prediction that Cel3A has beta-glucosidase activity based on sequence similarity to known β-glucosidases. In the presence of CMC, no cellulase alone was enough to confer growth, but all four enzymes expressed together yielded observable growth (Table 3.1.1).

Having observed that all four cellulases together conferred growth on CMC, we next asked whether all four are required to yield growth. To test which of the four enzymes were required for growth on CMC, cells individually expressing each YebF-cellulase fusion were co-inoculated in liquid cultures. Among these cultures, a strain expressing a particular cellulase was not inoculated to determine that cellulase was required for growth. Cultures lacking cells expressing YebF-Cel6A and/or YebF-Cel9A were able to grow, whereas cultures lacking cells expressing YebF-Cel3A or YebF-Cel5B were not able to grow (Table 3.1.2). To test whether cells simultaneously expressing YebF-Cel5B and YebF-Cel3A are also able to utilize CMC, we constructed plasmid pCEL2, which encodes YebF-Cel5B and YebF-CEl3A bicistronically. Wild-type cells harboring this plasmid were able to grow on CMC (Table 3.2.2). These data indicate that growth on CMC minimally requires beta-glucosidase activity and endoglucanase activity provided by Cel3A and Cel5B, respectively. No growth was observed in cells carrying pCEL4 in M9 minimial medium containing 1% AFEX-treated Avicel or 1% AFEX-treated corn stover (data not shown).

	Empty	Cel3A	Cel5B	Cel6A	Cel9A	CEL4
Glucose	+	+	+	+	+	+
Cellobiose	-	+	-	-	-	+
CMC	-	-	-	-	-	+

Table 3.1.1 Cell growth on glucose, cellobiose, or CMC as sole carbon sources. YebF-cellulase fusions (Cel3A, Cel5B, Cel6A and Cel9A) or all four cellulase fusions together (CEL4) were inoculated in M9 minimal medium containing 0.2% glucose, 0.34% cellobiose, or 1% CMC. (+) and (-) indicate growth or no growth as measured by culture turbidity after 48 hours.

	CEL4	CEL4 (-Cel3A)	CEL4 (-Cel5B)	CEL4 (-Cel6A)	CEL4 (-Cel9A)	CEL4 (-Cel6A, -Cel9A)	CEL2
CMC	+	-	-	+	+	+	+

Table 3.1.2 Cell growth on CMC as a sole carbon source. All four YebF-cellulase fusions (CEL4) or missing one cellulase fusion (-Cel3A, -Cel5B, -Cel6A, or -Cel9A) were co-inoculated in M9 minimal medium containing 1% CMC. (+) and (-) indicate growth or no growth as measured by culture turbidity after 48 hours.

Thus, far we have engineered normally non-cellulolytic *E. coli* cells with the ability to consume CMC as a sole carbon source by implementing the YebF secretion system for extracellular targeting of recombinant *C. japonicus* cellulases. The next step in the construction of a CBP platform is the conversion of the sugars obtained from cellulose hydrolysis to target chemical products. To advance to this step we chose to combine our engineered YebF-cellulases secretion system with a previously developed 1,2-propanediol pathway engineered for overexpression in *E. coli* (Altaras and Cameron, 1999). This pathway converts DHAP to 1,2-proanediol in three transformations mediated by the biosynthetic enzymes methyglyoxyl synthase (mgsA), 2,5-diketo-D-gluconic acid reductase (DkgA) and glycerol dehydrogenase (GldA) (all from *E. coli*) (Altaras and Cameron, 2000). These enzymes were cloned tri-

cistronically in expression vector pBAD18 to create plasmid p12PD and co-transformed with plasmid pCEL2 into wt W3110 cells. HPLC analysis revealed that these cultures produced 1,2-propanediol up to a concentration of 0.3 g/L. Cells lacking either plasmid p12PD and pCEL2 produced no measurable 1,2-PD, indicating that 1,2-PD production was dependent on both cellulolytic activity and overexpression of the 1,2-PD pathway.

3.2 A universal approach for engineering extracellular pathways

As a result of developing a CBP platform, we engineered cells capable of converting CMC to 1,2-PD by genetically amending *E. coli* cells with cellulase secretion activity and 1,2-PD overexpression. However, these cells showed approximately at 2% yield of 1,2-PD from CMC, while previous studies have shown that the same cells showed approximately an 8% yield of 1,2-PD from glucose (Altaras and Cameron, 2000). Thus, we chose to engineer the YebF pathway for enhanced secretion efficiency in part to increase downstream production yields in CBP platforms. However, we lacked a high-throughput screening platform to engineer the YebF secretion pathway. In fact, the general studying and engineering of extracellular secretion pathways has been limited by the lack of a universally available phenotype for screening large libraries to identify rare genetic conditions that affect secretion activity. Therefore, a high-throughput screening platform would be useful to better understand and engineer not just the YebF secretion pathway, but also other bacterial secretion systems used for secreting heterologous proteins, such as OsmY, OmpA, OmpF, and T2SS and T3SS.

To address this limitation, we developed a readily identifiable secretion phenotype that is compatible with automated high-throughput screening of large genetic libraries, and is universally amenable to study and engineer virtually any extracellular secretion system in

bacteria. We screened for a fluorescent phenotype created by the exogenous reagent FLASH-EDT2, which binds with high specificity to a tetracysteine motif CCXXCC (Griffin et al., 1998). The normally non-fluorescent FLASH-EDT2 becomes strongly fluorescent when bound to its ligand. Importantly, FLASH-EDT2 is not easily permeable to bacterial membranes (Ignatova and Gierasch, 2004), and could thus be added in situ to detect extracellular protein accumulation levels. We fused the C-terminus of YebF with peptide FLNCCPGCCMEP (labeled here as FT) that had previously been FACS-optimized for affinity and fluorescence, and followed by a 6H affinity tag (Fig. 3.2.1a) (Martin et al., 2005). Wild-type cells expressing the YebF-FT fusion protein generated a measurable fluorescent signal (Fig. 3.2.1b), and immunoblot analysis revealed a band corresponding to 6H in the supernatant (Fig. 3.2.1c). Interestingly, the signal was slightly enhanced in the absence of *dsbA*, presumably because the cysteine residues in the periplasmic compartment were maintained in a reduced state (Fig. 3.2.1b). As controls, YebF-FT lacking the Sec-dependent signal sequence (ΔssYebF-FT) or the FT epitope tag (YebF) generated no detectable signal (Fig. 3.2.1b), and no extracellular secretion was detected by immunoblot analysis (Fig. 3.2.1c). To test if YebF targeted for expression by the TAT-pathway could be secreted to the extracellular environment, the predicted Sec-dependent signal sequence was replaced with the leader sequence from native TAT substrate TorA (ssTorA- ΔssYebF). FlAsH and immunoblot analysis revealed that YebF-FT is not secreted to the extracellular medium when targeted for the Tat pathway, despite high level expression in the periplasmic compartment (Fig. 3.2.1b and c). These data indicate that YebF translocation to the culture medium is dependent on its mode of inner membrane secretion.

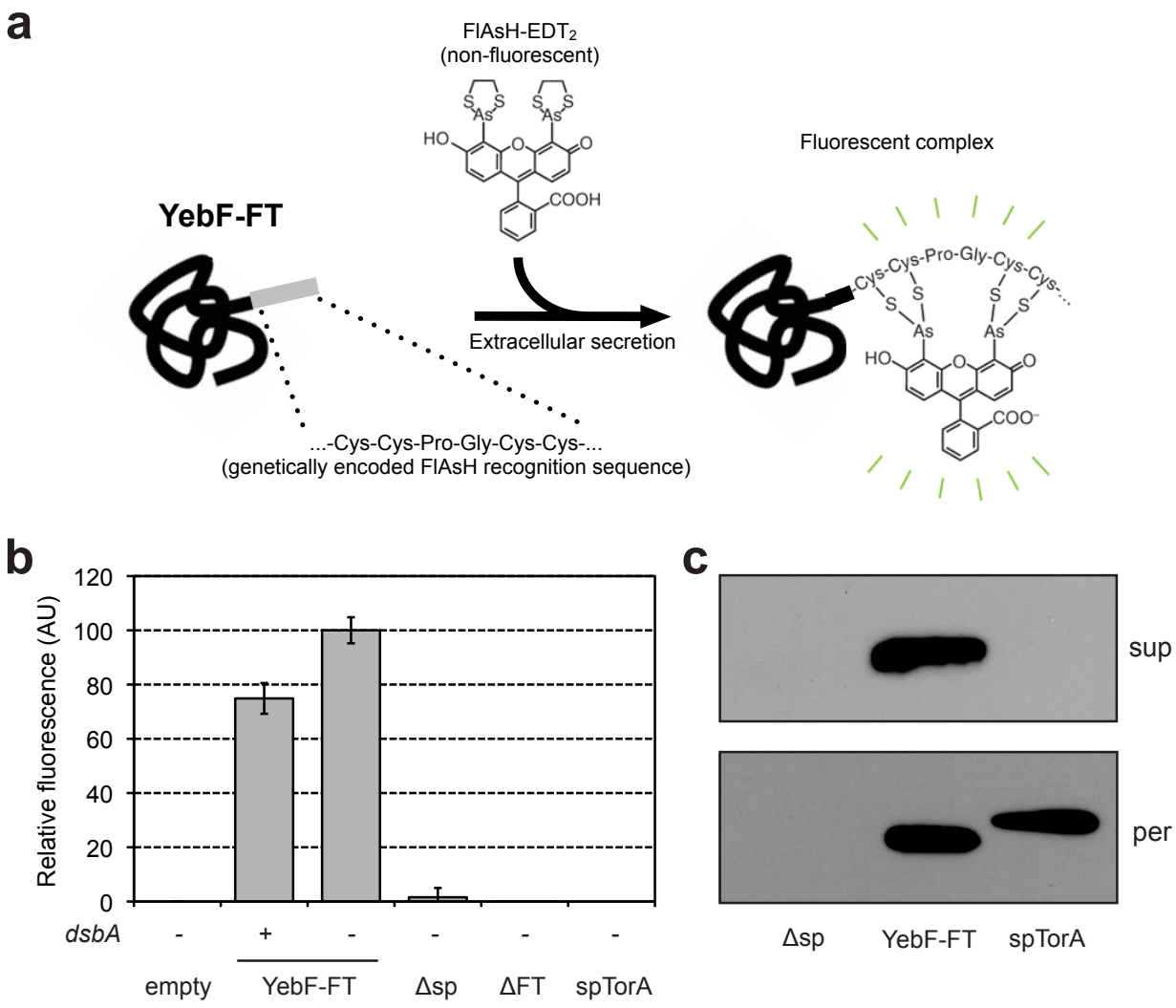

Figure 3.2.1 A high-throughput screening platform for extracellular secretion of proteins. (a) Schematic of YebF fused to tetracysteine epitope (FT) binding to FLASH-EDT2 creating a fluorescent complex. (b) Relative fluorescence measurements of wild-type and Δ*dsbA* cells expressing full-length YebF-FT and derivatives lacking the signal peptide (Δsp), lacking the FT (ΔFT), or with the signal peptide replaced with that of TorA (spTorA). (c) Western blot analysis of per and sup fractions of wild-type cells expressing YebF-FT and Δsp and spTorA derivatives. Detection of each was performed by using an anti-6H antibody.

The only genetic modification made for implementation of our FlAsH-based screen is the addition of a small epitope tag (18 amino acids) to the exoprotein of interest, in this case YebF, for specific interaction with the FlAsH-EDT2 reagent. Thus, we reasoned that this FlASH-based screening platform we developed initially for YebF could be transferred to other extracellular

secretion systems. To test this idea, we fused the C-terminal FT to other *E. coli* proteins that were previously observed to be released into the culture medium, OmpA (Xia et al., 2008), OmpF (Xia et al., 2008), and OsmY (Xia et al., 2008). Immediately following induction of these protein chimeras, the supernatant was isolated from the whole cell fraction to avoid labeling OmpA/F proteins that were retained within the outer membrane. FlAsH analysis of this cell-free supernatant resulted in measurable fluorescent signal for YebF-FT and OsmY-FT, and the YebF-FT signal was approximately twice as intense as that of OsmY-FT (Fig. 3.2.2a). No signal was observed from OmpA-FT and OmpF-FT (Fig. 3.2.2a). Immunoblot analysis corroborated the FlAsH-based observation that YebF-FT was secreted at significantly higher levels than OsmY-FT, and also that OmpA-FT, and OmpF-FT were not detectable in the supernatant (Fig. 3.2.2b). Therefore we concluded that OmpA-FT and OmpF-FT were not efficiently secreted under these conditions, possibly because they remain associated with the outer membrane. Collectively, these data indicate that the FlAsH-based assay for extracellular secretion is a versatile procedure that is amenable to not just one secretion system.

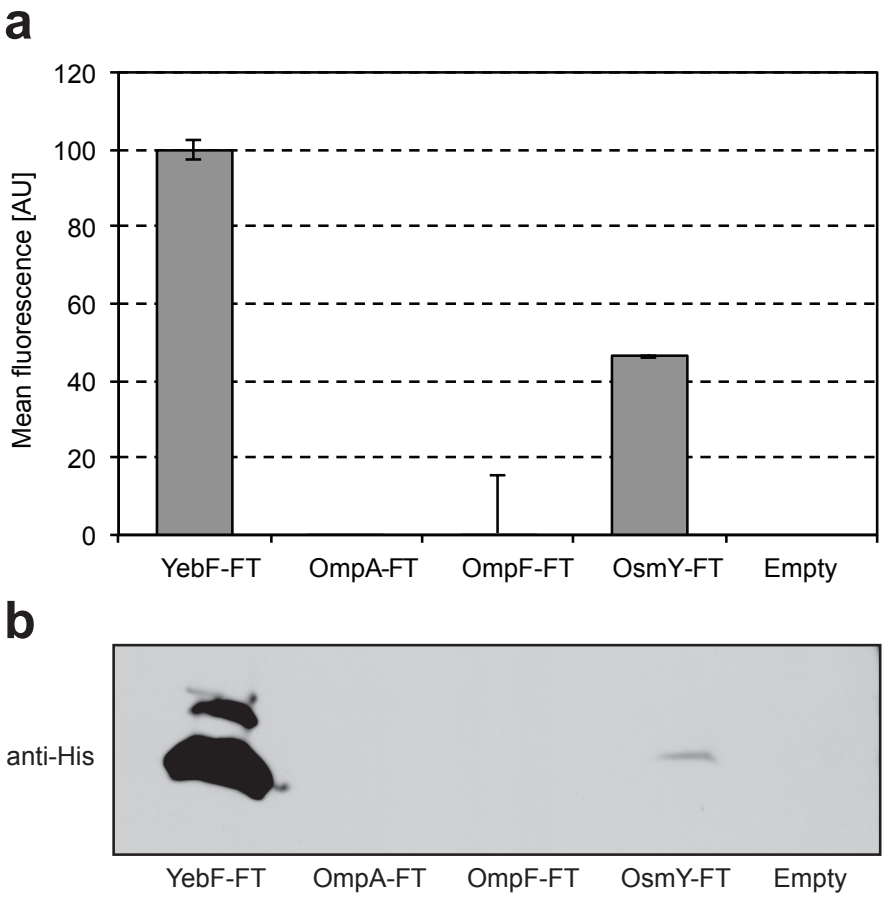

Figure 3.2.2 YebF is secreted at significantly higher levels than OmpA, OmpF, and OsmY. (a) FlAsH-analysis of C-terminal FT fusions in the cell-free supernatant. (b) Immunoblot analysis of each protein in the supernatant fraction. Detection of each was performed by using an anti-6H antibody.

Next, we tested if the FlAsH assay for protein secretion is compatible with T2SS. The T2SS encoded by the *out* gene cluster of *Dickeya dadantii* (formerly known as *Erwinia chrsyanthemi*) had previously been cloned in *E. coli* (He et al., 1991), enabling these cells to selectively secrete heterologous proteins, among them pectate lyase PelB. We fused the FT to the C-terminus of PelB and transformed this construct in wild-type MC4100 cells encoding the *out* gene cluster (out+). FlAsH analysis revealed a strong fluorescent signal in these cells over the same cells not carrying the *out* gene cluster (out-) (Fig. 3.2.3). Next, we removed genes encoding proteins essential for T2SS, *outC* and *outD* (Δ*outCD*) (Lindeberg et al., 1996) and observed no

secretion of PelB-FT, as expected (Fig. 3.2.3). Finally, we complemented the *outCD* genes with functional copies from *Dickeya dadantii* (Dd-OutC, Dd-OutCD) or closely related species *Pectobacterium carotovorum* (Pc-OutC, Pc-OutCD), on a separate plasmid. Only Dd-OutC and Dd-OutCD from *D. dadantii* were able to restore secretion of PelB-FT in Δ*outC* and Δ*outCD* cells, respectively. These data are consistent with the previous observation that *outC* and *outD* from *D. dadantii* cannot be complemented by *outC* and *outD* from *P. caratovorum*, whereas the other genes in the *D. dadantii out* cluster can be complemented by their *P. caratovorum* counterparts (Lindeberg et al., 1996). Taken together, these data suggest that *outCD* are 'gatekeepers' to species specificity of the T2SS. Importantly, the FlAsH assay developed here can be used in a high-throughput screen to target *P. carotovorum outCD* for laboratory evolution with the goal of overcoming the species specificity barrier of T2SS.

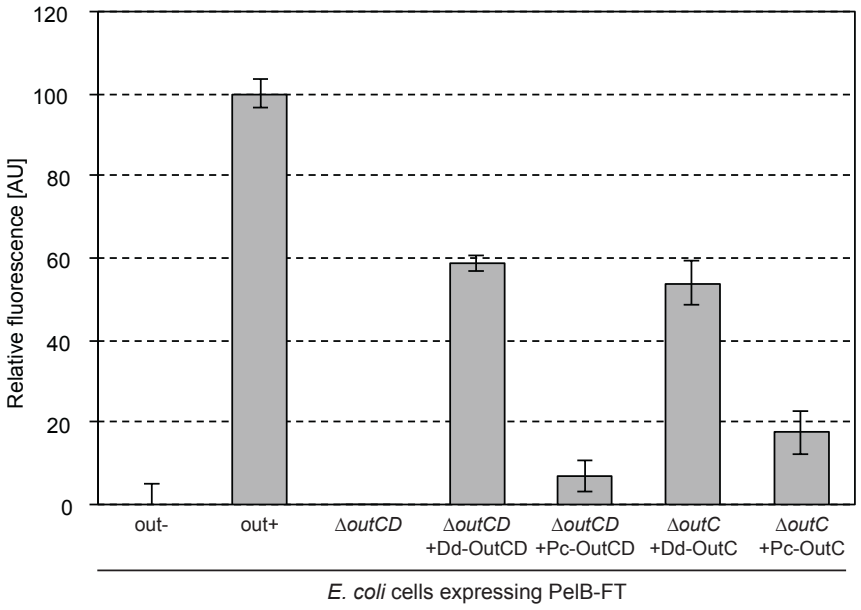

Figure 3.2.3 FlAsH analysis reveals species specificity among T2SS OutCD proteins. Wild-type *E. coli* cells were co-transformed with a plasmid encoding the *out* cluster from *D. dadantii* (out+) and a plasmid expressing the PelB-FT construct. Genes *outC* (Δ*outC*) or *outC* and *outD* together (Δ*outCD*) were removed, and functional copies from *D. dadantii* (Dd-OutC and Dd-OutCD) or from *P. carotovorum* (Pc_OutC and Pc_OutCD) were complemented *in trans*. These data confirm previously published results on *outCD* complementation (Lindeberg et al., 1996).

Finally, we tested if the FlAsH assay is compatible with the T3SS. The T3SS Hrp from *D. dadantii* functions in *E. coli* to secrete *Psuedomonas syringae* Avr proteins into plant cells and into the culture supernatant (Ham et al., 1998). Thus, we fused AvrPto with a C-terminal FT epitope and transformed this construct in DH5alpha cells carrying the *D. dadantii* Hrp system. Evidence from FlAsH analysis indicates that AvrPto-FT is secreted into the culture medium by these cells (Fig. 3.2.4). As a control, the same cells were used to express PelB-FT instead of AvrPto-FT, and this protein was not secreted at high levels, having only approximately 20% fluorescence signal compared to AvrPto-FT (Fig. 3.2.4a). Likewise, AvrPto-FT was not efficiently secreted by cells expressing the *D. dandantii* T2SS, and not the T3SS, while PelB-FT was secreted at high levels by these cells (Fig. 3.2.4a). Finally, we tested if the endogenous *E. coli* T2SS was amenable to FlAsH analysis. Thus, we fused ChiA to the C-terminal FT and transformed this construct into MC4100 cells lacking H-NS and over expressing the *gsp* gene cluster. These cells secreted detectable amounts of ChiA-FT, and were not able to secrete the AvrPto-FT protein (Fig. 3.2.4b). Interestingly, PelB-FT, which is a protein secreted by *D. dadantii* T2SS, was not secreted by the *E. coli* T2SS. These data indicate a stringent species specificity among gram-negative T2SS, which is consistent with previous observations (Lindeberg et al., 1996).

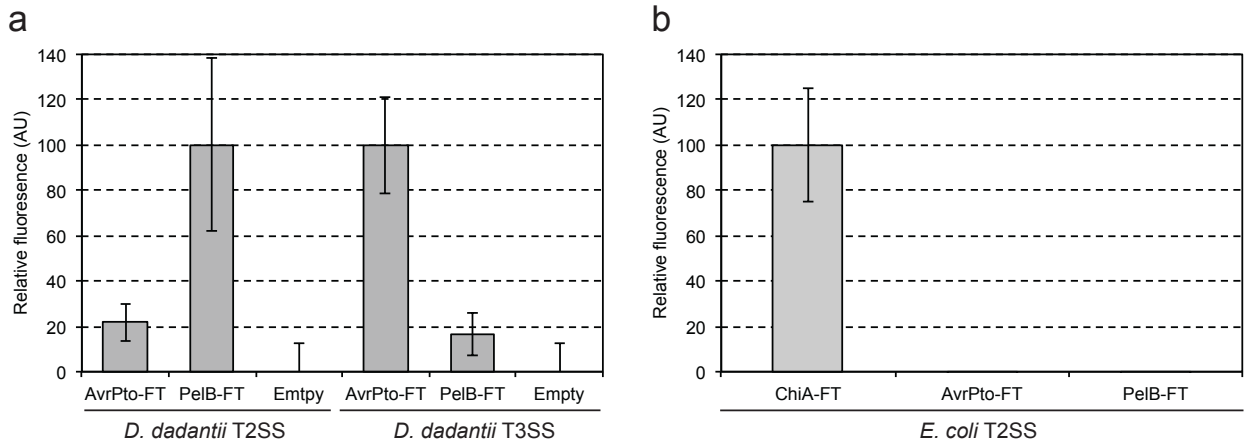

Figure 3.2.4 FlAsH analysis to measure secretion of T2SS and T3SS functioning in *E. coli*. MC4100 wt cells were transformed with *D. dadantii* T3SS and T2SS expression plasmids, and MC4100 Δ*hns* cells were transformed with *E. coli* T2SS expression plasmid. The secretion of AvrPto-FT, ChiA-FT, and PelB-FT in these strains was determined by FlAsH analysis.

3.3 Isolation of genetic mutations with enhanced YebF secretion efficiency.

Among the single-domain proteins studied so far, YebF yielded a comparatively robust secretion signal in FlAsH and immunoblot analysis, and was successful in secreting four recombinant cellulases. Therefore, we targeted YebF for screening a genome-wide Tn5 transposon insertion library to isolate genetic mutants that confer enhanced secretion activity. This transposon library was generated in BW21153 Δ*dsbA* cells carrying the YebF-FT expression plasmid (Donath et al., 2011). The resulting mutants were screened for fluorescence activity in the presence of FLASH-EDT2 and a B-score was assigned to each mutant as described previously (Malo et al., 2006). This analysis produced eight 'hits' with statistically significant increase in fluorescent signal (Fig. 3.3a). A 'clean' deletion of these genes in BW25113 (now *dsbA* positive) cells also resulted in improved the fluorescent signal over the parent strain (Fig. 3.3b). Finally, when functional copies of each gene was complement *in trans*, the fluorescence

activity returned to wild-type levels, indicating that each deletion was responsible for enhance secretion and that the phenotype is not due to polar effects (Fig. 3.3b).

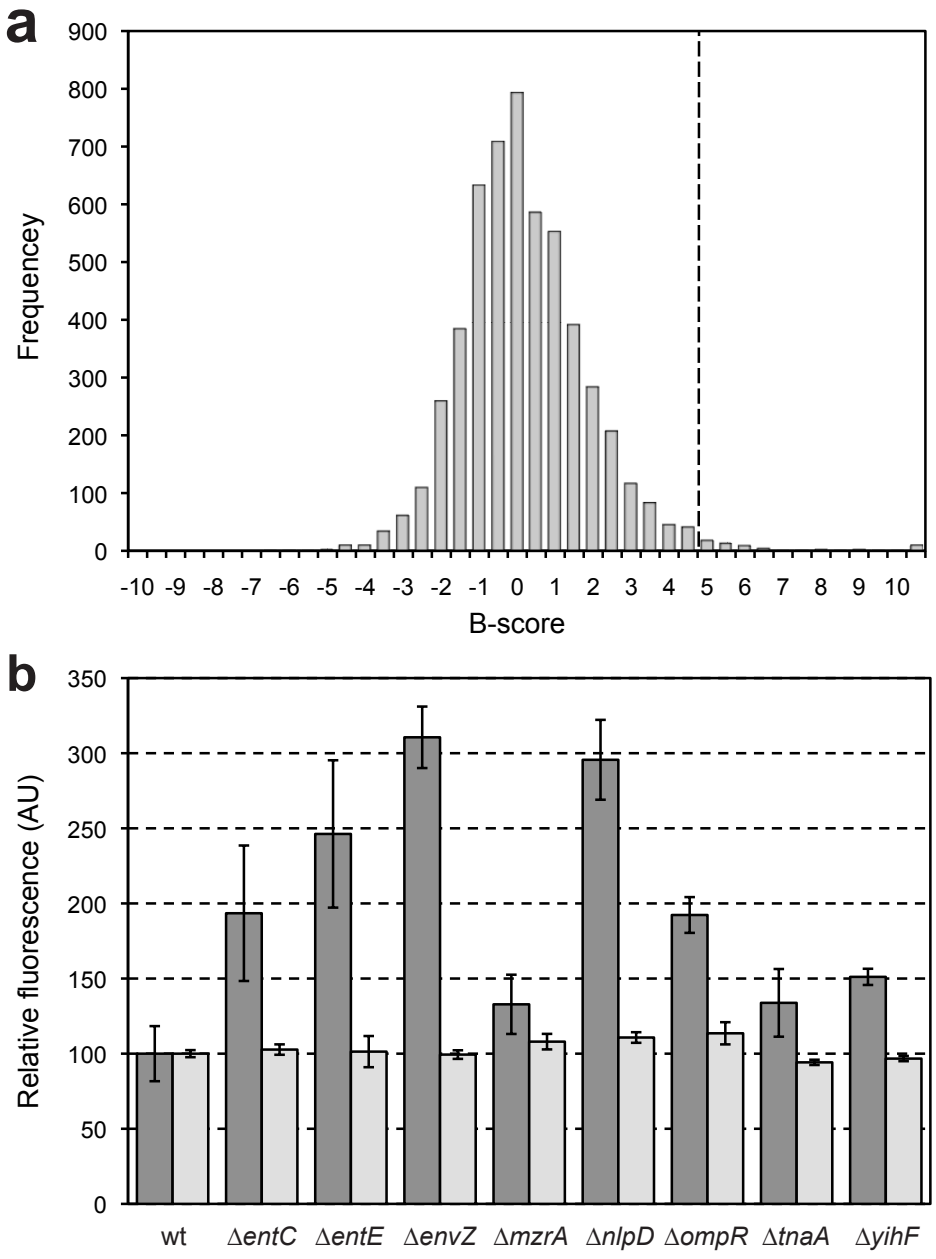

Figure 3.3 Isolation of genetic mutants with enhanced YebF transport. (*A*) B score analysis of 'hits' from high-throughput screening of transposon insertion library. (*B*) Fluorescent signals of cells with 'clean' deletions of genes identified in high-throughput screening (dark columns). Genes were complemented *in trans* (light columns).

3.4 Characterization of outer membrane permeability of isolated mutants.

The genetic mutations isolated here enhanced the extracellular accumulation of YebF. However, it is not clear if increased accumulation levels are the result of a compromise in membrane integrity in these mutants resulting in cellular proteins passively 'leaking' into the extracellular space, and not active secretion. To test this notion, we performed a thorough examination of membrane integrity by performing a series of previously established protocols that probe membrane integrity in *E. coli* mutants (McBroom et al., 2006). These protocols include measuring leakage of cellular proteins RNase I, MBP, and DsbA, and also sensitivity to the detergent deoxycholate (McBroom et al., 2006). The apparent hyper-secreting mutation Δ*ompR* exhibited a measurable amount of RNAse I, MBP, and DsbA leakage, as well as increased detergent sensitivity (Table 3.4). Importantly, these results are consistent with previous observations that *E. coli* Δ*ompR* cells have a compromised outer membrane and leak cellular proteins to the extracellular space (McBroom et al., 2006). Also, cells lacking *envZ* gave a similar phenotype in these analyses (Table 3.4). Based on these results, we concluded that increased secretion of YebF by cells lacking *envZ* or *ompR* may be at least partly attributed to increased outer membrane instability resulting in passive diffusion of cellular proteins into the extracellular space. For these reasons, cells lacking *envZ* or *ompR* were not included in further analyses.

Table 3.4 Characterization of outer membrane integrity and permeability.

	wt	Δ*entC*	Δ*entE*	Δ*envZ*	Δ*mzrA*	Δ*nlpD*	Δ*ompR*	Δ*tnaA*	Δ*yihF*
RNAseI leakage[a]	-	-	-	+	-	-	++	-	-
Detergent sensitivity[b]	-	+	-	++	-	+	++	-	-
MBP leakage[c]	-	-	-	++	+	-	+	-	-
DsbA leakage[c]	-	-	-	++	+	-	+	-	-

[a]-.no clearing; +, moderate clearing; ++, significant clearing.
[b]-, resistant; +, mildly sensitive; ++, sensitive.
[c]-, none detectable; +, low amount; ++, high amount.

3.5 Enhanced secretion of YebF-cellulase fusions

The mutations isolated in this study have significantly increased the extracellular accumulation of YebF. We were interested to know whether these mutations had the same effect on YebF-protein fusions, particularly the cellulase proteins that we used to develop a CBP platform earlier in this study. Thus, we next determined whether these mutations could upregulate the secretion of YebF-cellulase fusions. The same plasmids encoding the YebF-cellulase fusions (see Fig. 3.1) were transformed in wild-type and six mutant strains determined not to be leaky (see Table 3.4) and secretion of the fusion proteins to the supernatant fraction was detected by immunoblot analysis using an anti-6H antibody (Fig. 3.5a). As expected, each of the six mutations showed an increase in unfused YebF secretion relative to the wt host. Remarkably, all six mutants were able to dramatically upregulate the secretion of YebF-Cel3A into the supernatant, and all mutants, with the exception of Δ*entC* increased secretion of YebF-Cel5B (Fig. 3.5a)

If these mutants are more efficient at secreting YebF-Cel3A and YebF-Cel5B, the two enzymes required for growth on CMC (Fig. 3.1.2), then it stands to reason that these mutants have a faster growth rate in the presence of CMC when expressing these two proteins

simultaneously. Thus, we tested if these mutants had an increase in specific growth in the presence of CMC when expressing pCEL2. A growth curve was generated of cells carrying pCEL4 in M9 minimal medium containing 1% CMC and specific growth rates were calculated (Fig. 3.5b). Up to a 1.5 fold increase in growth rate was observed in Δ*tnaA* cells. We also tested whether these mutations increased 1,2-PD when introduced in our previously engineered CBP *E. coli* cells (Fig. 3.1), though no increase in 1,2-PD was observed (data not shown).

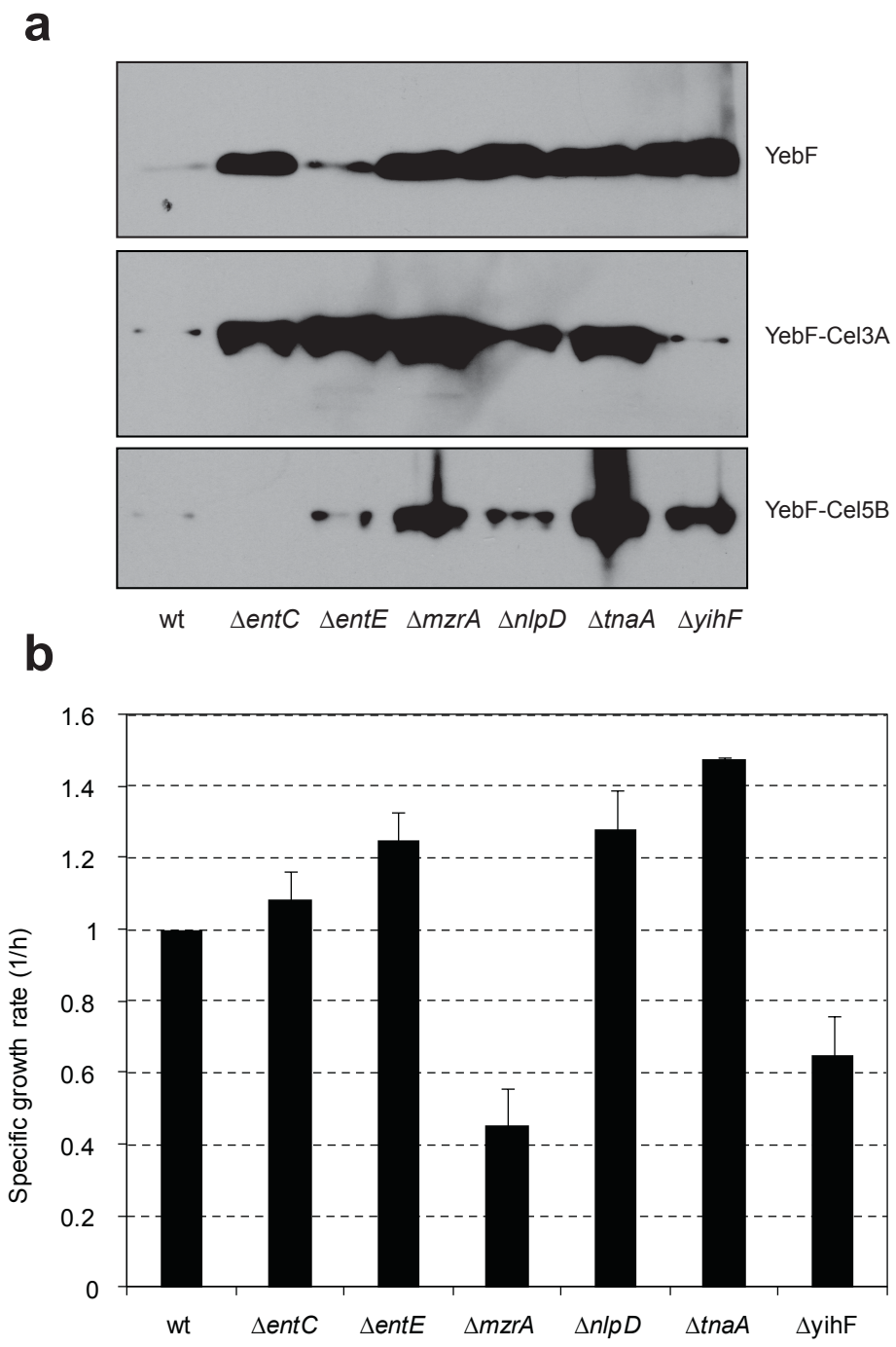

Figure 3.5 Enhanced extracellular secretion of YebF and YebF-cellulase fusions. (a) Extracellular secretion of YebF-cellulase fusions by hypersecretion mutations. Detection of YebF, YebF-Cel3A and YebF-Cel5B in cell-free supernatant fractions was performed using an anti-6H antibody. (b) Specific growth rate of hypersecreting mutants carrying plasmid pCEL2 in the presence of 0.34% cellobiose.

3.6 Discussion

This study and others have demonstrated that recombinant enzymes synthesized in *E. coli* can be targeted for export out of cells when fused C-terminally to the YebF protein (Zhang et al., 2006; Fisher et al., 2011). However, until now, the unanswered question remained whether the YebF pathway could be exploited to simultaneously co-secrete different heterologous proteins. This feature would be particularly useful in CBP organisms where more than one different cellulase is necessary to achieve sufficient hydrolysis of lignocellulosic substrates (Liao et al., 2011). Here, we show evidence that a single host can secrete at least four recombinant proteins when fused genetically to the YebF domain (Fig 3.1b). Furthermore, we discovered that of these four enzymes, only the predicted beta-glucosidase, encoded by Cel3A, and the endoglucanase, encoded by Cel5B, were required for the consumption of CMC as a sole carbon source. This observation was a critical step in developing a CBP platform that could convert the CMC to target chemical products. We chose the production of 1,2-propanediol because the metabolic pathway for 1,2-propanediol had previously been established and optimized in *E. coli* (Altaras and Cameron, 1999; 2000). Collectively, we demonstrated the direct conversion of CMC to measurable amounts of 1,2-propanediol. Our CBP system may be useful for engineering the CBP-mediated conversion of more complex cellulosic and lignocellulosic feedstocks, like pretreated switchgrass or corn stover.

The secretion of YebF into the periplasmic compartment occurs natively via the Sec pathway, and then YebF is translocated across the outermembrane by OmpF/C (Prehna et al., 2012) An interesting observation from our studies is that YebF targeted for expression via the Tat pathway is efficiently transported into the periplasm, but cannot be transported across the outer membrane. The YebF protein is likely in a folded state when it reaches the periplasm because the

intrinsic quality control mechanism of Tat transport dictates that secretory porteins are fully folded in the cytoplasmic compartment prior to export (DeLisa et al., 2003). Thus, one possible explanation is that YebF-mediated extracellular secretion requires an unfolded or semi-unfolded protein confirmation that may be required for threading through the lumen of OmpF (Prehna et al., 2012). This explanation is compatible with the ability of the YebF system to secrete different large recombinant proteins (Zhang et al., 2006; Fisher et al., 2011).

An important outcome of our studies is the development of a versatile assay for studying secretion systems by labeling secreted proteins with the FlAsH-EDT2 reagent. The reagent FlAsH-EDT2 was first developed to label intracellular proteins in eukaryotic cells (Griffin et al., 1998), but few studies have successfully demonstrated labeling intracellular proteins in bacteria, possibly due to the impermeability of the membrane to the FlAsH-EDT2 compound (Ignatova and Gierasch, 2004). While this may be a limitation for labeling proteins with FlAsH-EDT2 in certain applications, we relied on this impermeability to rapidly label secreted proteins in the extracellular environment without the need to first isolate the cell-free supernatant. This feature is particularly useful for automated screening of large genetic libraries. Thus, we were able to screen a genome-wide transposon insertion library, and we identified eight different genetic mutations that resulted in enhanced production of YebF in the culture medium. Remarkably, the enhanced accumulation of YebF in some of these mutants extended to YebF-protein fusions. This is an important observation because it represents a critical step in engineering superior expression hosts capable of secreting recombinant proteins to the culture medium at high titers.

3.7 Materials and methods

3.7.1 Bacterial strains, plasmids, and growth conditions

E. coli strain BW25113 (*lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$) and single-gene knockout mutants of BW25113 from the Keio collection (Baba et al., 2006) were used for all experiments unless otherwise noted. Strains were routinely grown aerobically at 37°C in Luria-Bertani (LB) medium, and antibiotics were supplemented at the following concentrations: ampicllin (Amp, 100 μg/ml); kanamycin (Kan, 50 μg/ml); chloramphenicol (Cm, 100 μg/ml). Protein synthesis was induced when the cells reached an absorbance at 600 nm (Abs600) of ~0.5 by adding 0.2% arabinose and/or 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG) to the media. *E. coli* strains were grown on M9 minimal medium (ref) supplemented with supplemented with MgSO4 (1 mM) and CaCl2 (0.1 mM). Glucose (0.2% [wt/vol]), cellobiose (0.34% [wt/vol]), carboxymethylcellulose (CMC) (1% [wt/vol]), and Avicel (1% [wt/vol]) were used as sole carbon and energy sources for cells grown in minimal medium.

3.7.2 Plasmid construction

Plasmid pTrc-YebF was created by insertion of DNA encoding yebF between SacI and XbaI of pTrc99a. Plasmid pCEL4 was synthesized (GeneArt) for codon-optimized polycistronic expression of cel6A, cel9A, cel5B, and cel3B. Plasmids pCel45A and pCel5D were synthesized (GeneArt) for codon-optimized expression of Cel45A and Cel5D, respectively. Each commercially synthesized cellulase was PCR amplified and inserted in between XbaI and HindIII of pTrc-YebF to create pTRC-YebF-Cel3A, pTRC-YebF-Cel5B, pTRC-YebF-Cel5D, pTRC-YebF-Cel6A, pTRC-YebF-Cel9A and pTRC-YebF-Cel45A, respectively. A reverse primer encoding 6X-His (6H) was used during PCR amplification to create a C-terminal 6H tag on each

plasmid. To create plasmid pYebF-Cel4, YebF-Cel6A, YebF-Cel9A, YebF-Cel5B, and YebF-Cel3B were PCR amplified with reverse primers containing 6H, HA, Flag, and Cmyc tags respectively. Then, each cellulase gene on pCEL4 was replaced with its respective YebF-cellulase fusion. To create pTrc-YebF-FT and pTrc-ΔspYebF-FT, YebF and ΔspYebF were PCR amplified using a reverse primer containing the FT and 6H sequence and inserted into SacI and XbaI of pTrc99a. Plasmids pTrc-spTorA-ΔspYebF-FT was created by inserting spTorA between NcoI and SacI of pTrc-ΔspYebF-FT. Plasmids expressing *entC*, *entE*, *envZ*, *mzrA*, *nlpD*, *ompR*. *tnaA*, and *yihH* were constructed by inserting DNA encoding a Shine-Dalgarno sequence followed by these genes in between SacI and XbaI of pBAD33. All primers were synthesized by Integrated DNA Technologies (IDT).

3.7.3 Protein analysis

Cells expressing recombinant proteins were harvested 4-6 h after induction at 30˚C. Cell-free supernatant fractions were prepared by centrifugation at 5,000 g for 10 min, and the proteins were precipitated overnight with 1% TCA at 4˚C. After centrifugation at 13,000 RPM for 30 min, the pellet was washed once with ice-cold acetone and dissolved in 1M Tris-HCl. pH 8.0 and an equal volume of electrophoresis loading buffer. The periplasm was separated from the cytoplasm using the osmotic shock method (Kim et al., 2005). All proteins samples were loaded in equal amounts for separation on an SDS-PAGE gel and then transferred to a nitrocellulose membrane for immunoblotting. The following primary antibodies were used: mouse anti-6H (1:3,000; Qiagen); mouse anti-FLAG (1:1,000; Stratagene); rabbit anti-C-myc (1:1,000; Sigma); rabbit anti-HA (1:1000; Sigma); rabbit anti-DsbA (1:20,000; gifted); mouse anti-MBP (1:2,000, New England Biolabs).

3.7.4 Transposon insertion library generation

A genome-wide transposon Tn5 insertion library was generated in strain EC100 using the EZ-Tn5™ <R6Kγori/KAN-2>Tnp Transposome Kit according to the manufacturer's instructions (Epicentre, Madison, WI). Following electroporation of the transposome into the recipient strain, the outgrowth was plated on LB agar medium supplemented with 50 μg/ml kanamycin. After overnight grown at 37°C, the plates were flooded with LB medium and any visible colonies scraped off with a cell spreader. The cell suspension was supplemented with glycerol to a final concentration of 30% glycerol and stored at -80°C. The Tn5 library frozen stock provided the donor strains for a P1-phage library of secretion mutants. Overnight cultures inoculated with the Tn5 library frozen stock were grown in 5 ml of LB medium supplemented with 50 μg/ml kanamycin. The overnight cultures were subcultured 1:25 in 50 ml of LB medium supplemented with 5 mM CaCl2, 0.2 % glucose, and 50 μg/ml kanamycin. Following 1 hour of incubation at 37°C with shaking, 0.625 ml of P1-phage stock was added to the flask. The growth of the culture was monitored by OD600. After several hours, a precipitous drop in OD600 was typically observed. The cultures were harvested 1 hour after the drop in OD600. Twenty-five ml of culture was added to two 50 ml conical centrifuge tubes along with 0.635 ml of chloroform. Each tube was vortexed briefly and centrifuged at 10,000 x g for 5 minutes. The resulting P1 lysates from the supernatant were stored at 4°C and used to produce YebF secretion libraries. Thirty milliliters of overnight culture (30 ml) of *E. coli* BW25113 ΔdsbA harboring pTrc-YebF-FT was centrifuged at 6,000 xg for 5 min. The pellet was resuspended in 30 mL of LB medium supplemented with 100 mM MgSO4 and 5 mM CaCl2 then incubated at 30°C with 3 ml of P1 phage stock. After 30 minutes, 6 ml of 1 M sodium citrate was added to stop the reaction. The

cells were again pelleted at 6,000 x g for 5 minutes and resuspended in 10 mL of LB supplemented with 100 mM sodium citrate. The transduced cells were plated onto Omniwell plates (Nunc) containing LB agar medium supplemented with 50 μg/ml kanamycin and 100 μg/ml ampicillin.

3.7.5 FlAsH Screening of YebF secretion mutants

Individual clones were robotically picked into wells of 96-well bar-coded microtiter plates containing 100 μl of LB medium supplemented with 50 μg/ml kanamycin and 100 μg/ml ampicillin. The picked clones were incubated with shaking overnight at 30°C and high humidity. The plates containing overnight cultures were loaded onto a Tecan Freedom EVO 200 liquid handling robot where all remaining processing was performed. The overnight cultures were subcultured (1:10) into clear-bottom black microtiter plates with 180 μl of LB medium in each well supplemented with 0.5 mM IPTG, 50 μg/ml kanamycin, and 100 μg/ml ampicillin. The remaining overnight culture was supplemented with glycerol to a final concentration of 30% glycerol and stored at -80°C. After 24 hours incubation with shaking at 30°C and high humidity, the subcultured plates were assayed for Flash-tagged YebF secretion. Each culture was assayed for cell growth by measuring OD600, then 20 μl of concentrated assay cocktail containing 10 μM FlAsH-EDT (Invitrogen) and 500 μM 2,3 dimercaptopropanol was added to each well. The plates incubated with shaking for an additional hour before reading the culture fluorescence 485 nm/528 nm.

3.7.6 Characterization of outer membrane permeability

Assays for detergent sensitivity and RNase I, MBP, and DsbA leakage were performed essentially as described before (McBroom et al., 2006).

CHAPTER 4

AN ENGINEERED GENETIC SELECTION FOR EXTRACELLULAR SECRETION

4.0 Introduction

The approach for engineering extracellular secretion pathways developed in Chapter 3 relies on labeling exoproteins with an exogenous reagent FlAsH-EDT2. The nature of this assay means that genetic or other libraries must be screened by a mechanical system, such as with a 96- or 384-well plate reader. Even with automated systems, this can limit the size of the libraries screened. Therefore, this approach may not be fruitful for some applications that require large library sizes or multiple combinatorial libraries, such as the laboratory evolution of proteins (Stemmer, 1994; You and Arnold, 1996). An alternative approach where extracellular protein secretion is linked to cell viability would streamline the isolation of rare genetic events that affect secretion activity by eliminating the need to screen individual clones. For example, a genetic selection where the non-secretory phenotype is lethal would be more suitable for laboratory evolution of protein machinery and screening large or multiple genetic libraries.

Beta-lactam antibiotics are among the most frequently used antibiotics, and the activity of beta-lactamase (Bla) enzymes has been used extensively in engineered genetic selections for studying and engineering protein secretion, solubility, and stability (Fisher et al., 2006; Waraho and DeLisa, 2009; Mansell et al., 2010). One method for overcoming beta-lactamase-mediated drug resistance is by a 165-amino acid beta-lactamase inhibitor protein (BLIP) that is synthesized by *Streptomyces clavuligerus* (Paradkar et al., 1996). The BLIP protein is a potent inhibitor of class A beta-lactamase, including TEM1 (Ki = 0.1 to 0.6 nm) (Zhang and Palzkill,

2003). This activity of BLIP has previously been harnessed in a dual genetic selection to screen error-prone libraries of LuxR for altered signaling specificity in quorum sensing (Collins et al., 2006). Here, we describe the use of BLIP / Bla in an engineered genetic selection for studying and engineering extracellular secretion pathways. This strategy was successfully implemented for studying the secretory mechanisms of the YebF pathway and also the type II secretion system.

4.1 An engineered genetic selection for extracellular secretion of proteins

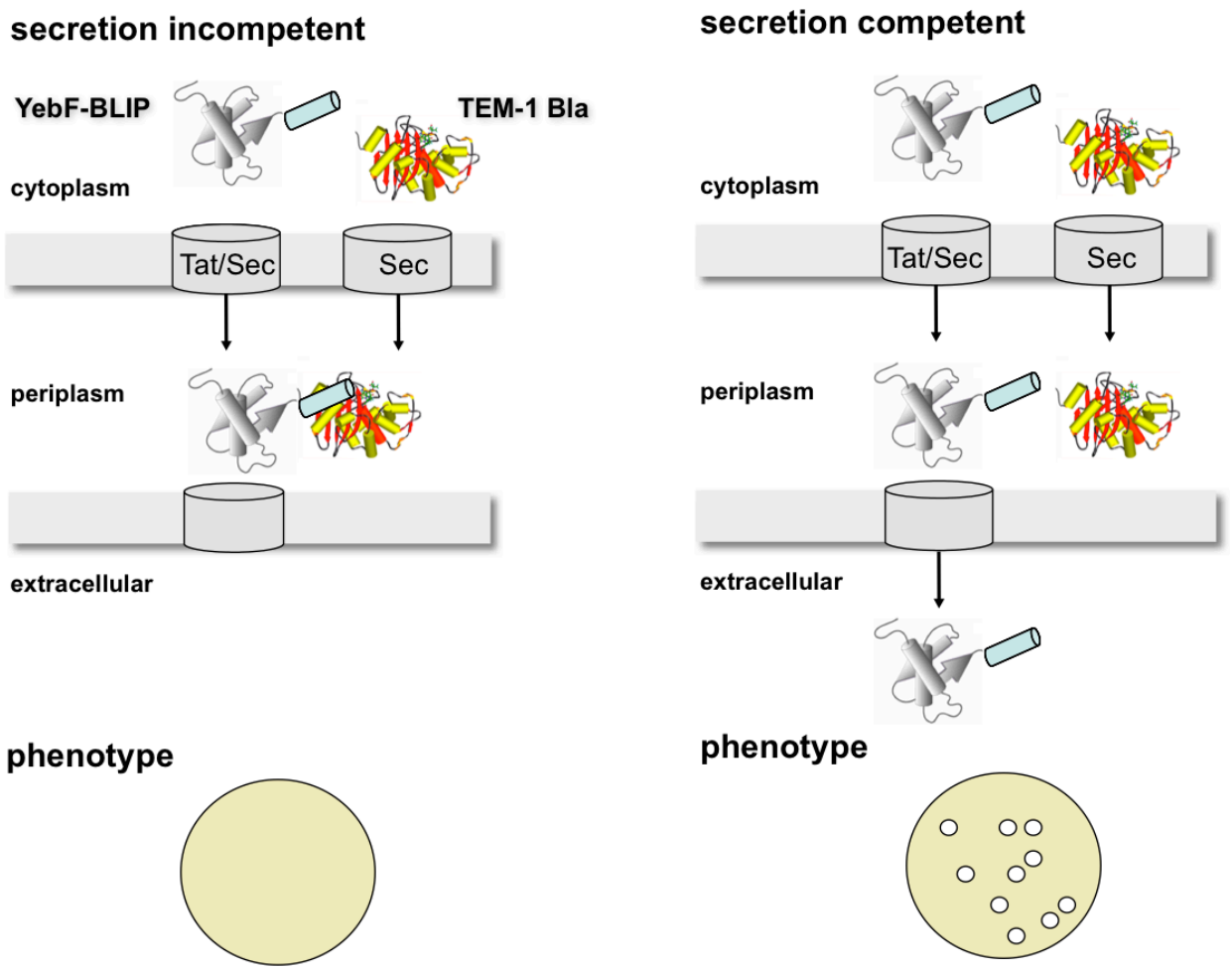

Figure 4.1.1 An engineered genetic selection for extracellular secretion of proteins. The exoprotein of interest, in this case YebF, is fused N-terminally to BLIP and TEM-1 Bla is co-expressed in these cells. Under conditions that hinder the translocation of YebF-BLIP across the outer membrane, the YebF-BLIP protein is retained in the periplasm where it inhibits the activity of Bla, and the result is that cells are less viable in the presence of beta-lactam antibiotics (left panel). By contrast, under conditions of secretion, the YebF-BLIP protein is localized to the extracellular space and Bla remains active, and the result is that cells are more resistant to beta-lactam antibiotics (right panel).

To create a genetic selection for extracellular secretion, we created a host-vector system where the non-secretory phenotype is lethal. We relied on the beta-lactamase inhibitor protein (BLIP), which binds to and inhibits β-lactamase (Bla) with a Ki of 0.1 to 0.6 nM (Zhang and

Palzkill, 2003). We genetically fused BLIP to the C-terminus of YebF and hypothesized that cells would become less viable under conditions of non-secretion in the presence of beta-lactam antibiotics, such as carbenicillin (Carb) (Fig. 4.1.1). A leaderless YebF domain was inserted between the Sec-dependent signal sequence of BLIP (ssBLIP) and mature protein of BLIP (dssBLIP) to create a YebF-BLIP fusion with the native BLIP signal sequence. This genetic fusion was ligated into expression vector pTrc99a (Amann et al., 1988), whereby Bla is expressed constitutively from the vector backbone, creating plasmid pTRC-YebF-BLIP. As controls, we made derivatives of pTRC-YebF-BLIP that express full-length BLIP without the YebF domain (pTRC-BLIP), and also YebF-BLIP without an inner membrane secretion signal (pTRC-Δss-YebF-BLIP). These plasmids were tranformed in wt BW25113 cells and overnight cultures of were diluted 10^{-1} to 10^{-8} and each dilution was spotted in equal volumes on LB agar plates containing a range of Carb concentrations from 0 (-Carb) to 100 μg/mLμg/mL (+Carb) (Fig. 4.1.2a). In this spot plate analysis, we observed that the expression BLIP protein without the YebF fusion was significantly attenuated for growth compared to the YebF-BLIP fusion (Fig. 4.1.2a). As expected, cytoplasmic expression of the YebF-BLIP construct encoded by ΔssYebF-BLIP showed no observable attenuation of growth compared to empty vector cells. To test if the difference in cell viability between YebF-BLIP and BLIP was due to the localization of the BLIP protein, we performed an immunoblot analysis of cell-free supernatant fractions from cells expressing each protein. In cells expresing YebF-BLIP, a strong band corresponding YebF-BLIP, whereas only a faint band corresponding to BLIP appeared in cells expressing the BLIP construct (Fig. 4.1.2b) Finally, no band was detected in cells expressing Δss-YebF-BLIP. These results indicate that efficient secretion of BLIP requires the presence of the N-terminal signal sequence

and the YebF secretion domain. It also supports the notion that cell viability in spot plating analysis is linked to the localization of the BLIP protein.

Encouraged by these results, we tested if we could select for cells expressing the YebF-BLIP fusion reporter from a mixed population of cells expressing either YebF-BLIP or BLIP proteins. Overnight cultures of the YebF-BLIP and BLIP hosts were mixed in equal amounts and plated to single colony resolution on 100 μg/mL Carb and 0 μg/mL Carb. Plasmids from ten randomly selected colonies were isolated and treated with EcoRI and HindIII to release either the BLIP or YebF-BLIP inserts, with predicted sizes of 672 bp and 948 bp, respectively. These studies revealed that 10/10 selected colonies from the Carb (100 μg/mL) plates had plasmids with the YebF-BLIP insert (data not shown). Only 3/10 colonies selected in the absence of Carb had the YebF-BLIP insert (data not shown). These data indicate that we can effectively select against cells not efficiently secreting the BLIP protein.

Finally, we were interested to know if this BLIP / Bla genetic selection was limited to just the YebF pathway, or if it could be extended to other secretion systems. To test if we could generate a similar genetic selection for T2SS, we fused the T2SS-secreted protein PelB to the N-terminus of BLIP followed by a 6H affinity tag and cloned this fragment into expression vector pTrc99a (Amann et al., 1988) for constitutive co-expression of Bla from the vector backbone. This plasmid was transformed in cells expressing the functional *out* cluster for *Dickeya dadantii,* and also the nonfunctional *out* cluster lacking the essential genes *outC* and *outD*. In cells encoding non-fuctional *out* clusters, functional copies of *outC* and *outD* from *D. dadantii* and *P. carotovorum* were provided on a separate plasmid. These cultures were diluted 10^{-1} and plated on LB agar containing 25 μg/mL Carb (Fig. 4.1.2). Under these conditions, only cells containing *outCD* genes from *D. dadantii* grew, where as cells expressing *outCD* from *P. carotovorum* did

not grow (Fig. 4.1.2c). This is consistent with our previous observation that PelB is secreted by *D. dadantii outCD* but not by *P. carotovorum outCD* (see Fig. 3.2.3), and this is also consistent with a previous study on this T2SS (Lindeberg et al., 1996). Taken together, these data provide compelling preliminary evidence that the BLIP / Bla selection can be used to study and engineer T2SS and YebF secretion pathways.

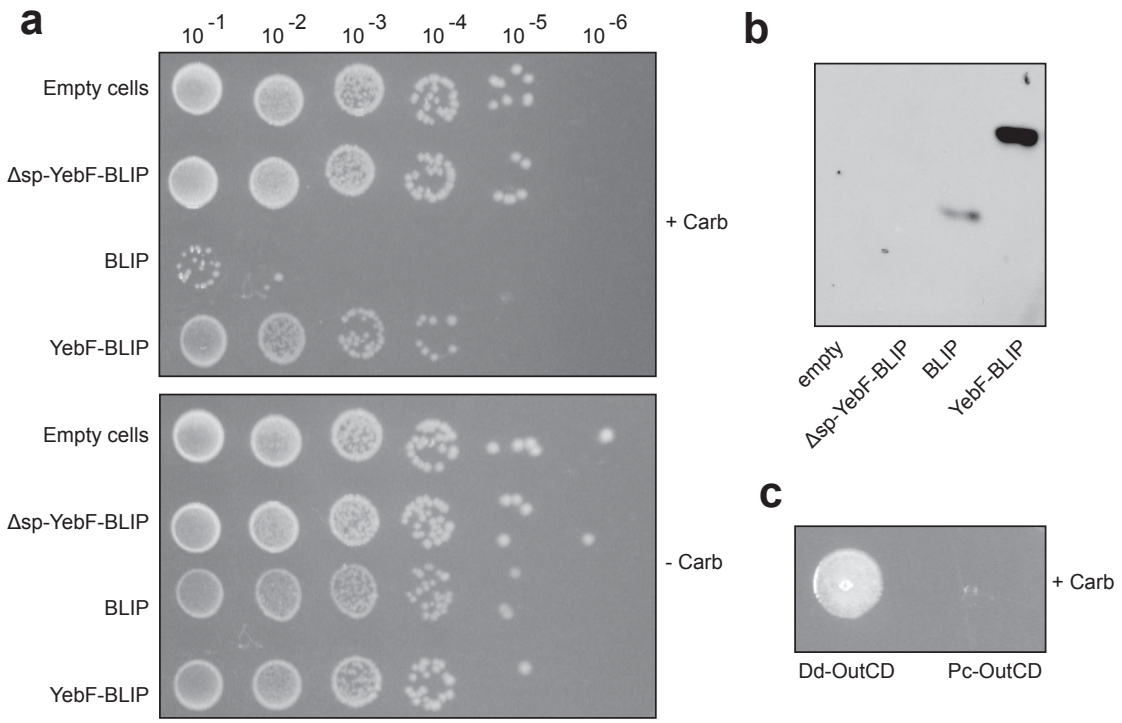

Figure 4.1.2 A BLIP / Bla genetic selection for engineering YebF and T2SS pathways (a) Wild-type cells harboring an empty vector (empty), a YebF-BLIP fusion lacking the Sec-dependent signal sequence (Δss-YebF-BLIP), full-length blip lacking the YebF secretion domain (BLIP), and a full-length YebF-BLIP fusion (YebF-BLIP) were diluted 1/10 and spotted in equal amounts on 100 μg/mL Carb (+ Carb) or no Carb (- Carb). (b) Immunoblot analysis of Δss-YebF-BLIP, BLIP, and YebF-BLIP proteins. Proteins were detected in the supernatant fraction using an anti-6H antibody. (c) Overnight cultures of MC4100 cells harboring PelB-BLIP and either the *out* gene cluster (out+) or the *out* gene cluster lack *outCD* (Δ*outCD*) were diluted 1/10 and spot plated on 25 μg/mL Carb (+Carb).

4.2 Discussion

To our knowledge, this study provides the first evidence for a general genetic selection for extracellular protein secretion in bacteria. The versatility of this genetic selection is in part attributed to the BLIP protein, which is a small protein (167 aa) that can be fused to virtually any exoprotein of interest. In this manner, the BLIP / Bla genetic selection is not limited to just one extracellular secretory mechanism, but can be in principle be applied to any two-step extracellular secretion system in gram-negative bacteria. This speculation is supported by our observation that the genetic selection was compatible with both the YebF secretion pathway and also the T2SS from *D. dadantii* functioning in *E. coli*.

We predict that this selection technology will be useful for reprogramming secretory pathways with enhanced efficiency and new activities. For instance, it should be possible to screen variants of YebF or other proteins in the YebF pathway, such OmpF/C or OmpX, to find mutations that enhance the secretion activity of YebF or YebF-protein fusions. Also, the T2SS is known to have a stringent species specificity whereby proteins from a closely related organism cannot be secreted by the secretion system of another organism (Lindeberg et al., 1996). This species specificity barrier is thought to be encoded by the *outC* and *outD* genes as they are the only genes in the *D. dadantii out* cluster that cannot be complemented by those from the closely related organism *P. carotovorum* (Lindeberg et al., 1996). Importantly, these results are consistent with our observation that cells functioning in our selection system are significantly less viable when expressing the Pc-OutCD instead of the Dd-OutCD. Thus, our system may be well-suited for laboratory evolution of the Pc-OutCD proteins to find variants with relaxed substrate specificity.

4.3 Materials and methods

4.3.1 Bacterial strains, plasmids, and growth conditions

E. coli strain BW25113 (*lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$) were used for all experiments unless otherwise noted. Strains were routinely grown aerobically at 37°C in Luria-Bertani (LB) medium, and antibiotics were supplemented at the following concentrations: ampicllin (Amp, 100 μg/ml); kanamycin (Kan, 50 μg/ml); chloramphenicol (Cm, 100 μg/ml); spectinomycin (Spec, 100 μg/mL). Protein synthesis was induced when the cells reached an absorbance at 600 nm (Abs600) of ~0.5 by adding 0.2% arabinose and/or 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG) to the media.

4.7.2 Plasmid construction

Plasmid pTRC-ssBLIP-YebF-BLIP was created by inserting the ssBLIP signal sequence in EcoRI SacI, leaderless YebF in SacI XbaI and leaderless BLIP with a C-terminal 6H affinity tag between XbaI and HindIII of pTrc99a. Plasmid pTRC-dssYebF-BLIP and pTRC-BLIP were created by removing the ssBLIP signal sequence and leaderless YebF by restriction digest, respectively. Plasmid pTRC-PelB-BLIP was created by inserting PelB between SacI and XbaI and leaderless BLIP with a C-terminal 6H between XbaI and HindIII. The sequence of all plasmids was confirmed by sequencing.

4.7.3 Protein analysis

Cells expressing recombinant proteins were harvested 4-6 h after induction at 30˚C. Cell-free supernatant fractions were prepared by centrifugation at 5,000 g for 10 min, and the proteins were precipitated overnight with 1% TCA at 4˚C. After centrifugation at 13,000 RPM for 30 min,

the pellet was washed once with ice-cold acetone and dissolved in 1M Tris-HCl. pH 8.0 and an equal volume of electrophoresis loading buffer. All proteins samples were loaded in equal amounts for separation on an SDS-PAGE gel and then transferred to a nitrocellulose membrane for immunoblotting. The following primary antibodies were used: HRP-conjugated mouse anti-6H (1:3,000; Qiagen).

CHAPTER 5

ENVELOPE STRESS IS A TRIGGER FOR CRISPR RNA-MEDIATED DNA SILENCING IN *ESCHERICHIA COLI*[2]

The ability to move proteins across the cytoplasmic membrane, and very often out of the cell completely, is a key aspect of the physiology and pathogenicity of gram-negative bacteria. Thus, the exploitation of cellular protein biosynthesis machinery and secretory mechanisms is not without consequences to the host. Protein overexpression can induce cell stress responses, particularly when these proteins misfold and are targeted for secretion pathways (Darwin, 2005). Because protein secretion is essential for cellular physiology, 'jamming' secretion pathways with misfolded proteins can induce envelope stress responses that can regulate genes for maintaining cell envelope homeostasis. For example, in *E. coli* the mislocalization of secreted proteins that disrupt the proton motive force along the inner membrane induces stress response pathway called the phage-shock response (Psp) (Darwin, 2005). This observation is supported by a related study that showed that saturation due to protein overexpression along the Tat pathway, which relies on the proton motive force for energy, can be relieved by the Psp protein PspA (DeLisa et al., 2004). Protein misfolding can also induce additional stress response pathways with different but

---

[2]Adapted with permission from:

Perez-Rodriguez R.*, **C.H. Haitjema***, Q. Huang, K. H. Nam, S. Bernardis, A. Ke, and M. P. DeLisa. 2011. Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Mol Microbiol* **79**:584–599. *These authors contributed equally to this work

Nam K.H., **C.H. Haitjema**, F. Ding, M.P. DeLisa, and A. Ke. 2012. Cas5d protein processes pre-crRNA and assembles into a Cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* (*in press*)

overlapping functions. The Bae and Cpx two-component system response regulators have been known to respond to envelope stress caused by protein misfolding (MacRitchie et al., 2008).

In this chapter, I describe the Bae-mediated response to a misfolded protein targeted for secretion by the Tat pathway. Interestingly, part of this Bae-mediatesd response is the induction of the CRISPR Cas system, which functions to silence the plasmid DNA encoding the misfolded protein. It should be noted that much of the work presented in this chapter is by a co-author, Ritsdeliz Perez-Rodriguez. While I am also a co-first author of this work, my contribution to this project are in the latter parts of this chapter, particularly subsections 5.5 through 5.7. The other sections are included to provide context. Finally, section 5.8 is a separate but related study of mine on the CRISPR Cas Dvulg subtype, encoded by *Bacillus halodurans,* which was partially reconstituted in *E. coli* cells to reveal functional complementarity between these two systems.

5.0 Introduction

The use of small RNAs to regulate gene expression is ubiquitous in all living organisms (Waters and Storz, 2009). In one remarkable instance, bacteria and archaea acquire resistance to bacteriophages and conjugative plasmids by employing an RNA-mediated defense mechanism against these foreign invaders. In this process, short fragments (~24–48 nucleotides) of the invading DNA are integrated in the genome as spacers between similarly sized clusters of regularly interspaced short palindromic repeats (CRISPR) loci (Bolotin et al., 2004; Makarova et al., 2006; Barrangou et al., 2007; Brouns et al., 2008; Sorek et al., 2008; Lillestøl et al., 2009). CRISPR loci have been identified in nearly all archaeal genomes and almost half of eubacterial genomes that have been sequenced to date (Sorek et al., 2008). They are often adjacent to an operon that encodes the CRISPR-associated (Cas) proteins (Fig. 5.1A), which are predicted

RNA binding proteins, endo- and exo-nucleases, helicases and polymerases (Haft et al., 2005; Makarova et al., 2006). Bacteria encoding CRISPR-Cas systems that have been exposed to a virus or conjugative plasmid integrate phage- or plasmid-derived spacer sequences into the leader-proximal end of their CRISPR loci (Barrangou et al., 2007). The appearance of acquired spacer DNA in the genomes of these bacteria correlates with viral or plasmid resistance in a process that is dependent upon several of the *cas* gene products (Barrangou et al., 2007; Brouns et al., 2008).

A key feature of CRISPR-encoded immunity is an inhibitory ribonucleoprotein complex comprised of a subset of Cas proteins and a guide RNA. This complex is believed to be responsible for targeting foreign genetic elements through base pairing between the bound RNA guide and either the sense or antisense strands of the target (Brouns et al., 2008). The formation of the guide RNAs begins when the CRISPR repeat-spacer arrays are transcribed from the leader region, producing a CRISPR transcript called the pre-crRNA (Hale et al., 2008; Lillestøl et al., 2009). The full-length pre-crRNAs are subsequently processed into small CRISPR RNA (crRNA) molecules that correspond to a spacer flanked by two partial repeats (Tang et al., 2002; Brouns et al., 2008; Hale et al., 2008). How the expression of CRISPR arrays and *cas* genes is regulated when the cell is threatened by foreign DNA is poorly understood, although in *Escherichia coli* it appears to involve H-NS repression, which silences transcription from the CRISPR-Cas promoters (Pul et al., 2010), unless relieved by the transcriptional activator LeuO (Westra et al., 2010).

The production of crRNAs is carried out by distinct Cas proteins in different organisms. In *Pyrococcus furiosus*, Cas6 processes pre-crRNA to release individual invader-targeting RNAs and remains bound to the CRISPR repeat sequences at the 5′ end of the cleavage product (Carte

et al., 2008). In *E. coli*, which lack Cas6, as many as five Cas proteins – CasABCDE – reportedly form a complex called CRISPR-associated complex for antiviral defence (Cascade) that processes CRISPR RNA (Brouns et al., 2008). Of these five enzymes, the catalytic activity of the CasE subunit is essential for pre-crRNA cleavage, which occurs within each repeat of the CRISPR RNA precursor (Brouns et al., 2008; Carte et al., 2008). Processed crRNAs encode the entire spacer unit flanked by the last eight bases of the repeat sequence at the 5′ end and a less well-defined 3′ sequence that ends in the next repeat region (Brouns et al., 2008; Carte et al., 2008). As mature crRNAs are similar at their 5′ but not their 3′ ends implies that this region may serve as a conserved binding site for Cas effector subunits, as suggested previously (Kunin et al., 2007; Carte et al., 2008). However, whether Cas6 or CasE is associated with the crRNA-Cas ribonucleoprotein (crRNP) effector complex, thereby coupling guide RNA biogenesis with target degradation, is currently not known. In fact, there are very few details about the identity and mechanism of the crRNP complex. Finally, while recent observations in *E. coli* and *Staphylococcus epidermidis* implicate DNA as the CRISPR-Cas target (Brouns et al., 2008; Marraffini and Sontheimer, 2008), there is little direct evidence to support this. In fact, in *P. furiosus* the target appears to be RNA (Hale et al., 2009).

Here, we show that the *E. coli* CRISPR-Cas system is triggered by the expression of a plasmid-encoded protein targeted for export by the twin-arginine translocation (Tat) pathway in cells that lack the DnaK molecular chaperone. Expression of this Tat substrate activated the CRISPR-Cas system, which resulted in specific silencing of the plasmid DNA that encoded the substrate. This silencing required: (i) base complementarity between genomic spacer DNA and the plasmid, (ii) specific targeting of the substrate to the Tat pathway, (iii) the BaeSR two-component signal transduction system and (iv) each of the Cas enzymes (CasABCDE and Cas1–

3). At the heart of this silencing is the CasE protein, which formed a stable ternary complex with CasC and CasD, and functioned as the catalytic core of the Cas system to process pre-CRISPR RNA into mature crRNAs.

5.1 Identification of CasE as a key component in prokaryotic gene silencing

Previously, we observed that a reporter protein consisting of the Tat signal peptide derived from *E. coli* trimethylamine *N*-oxide reductase (ssTorA) fused to the green fluorescent protein (GFP) was undetectable following its expression in *E. coli* cells that lacked the molecular chaperone DnaK (Fig. 5.1.1B and C) (Pérez-Rodríguez et al., 2007). We initially suspected that the ssTorA–GFP protein was unstable in the absence of DnaK, rendering it susceptible to proteolytic degradation. However, Northern analysis revealed that the absence of ssTorA–GFP in *E. coli* BW25113 Δ*dnaK* cells was due to elimination of the encoding mRNA (Fig. 5.1.1D). When plasmid-encoded DnaK was provided *in trans* to Δ*dnaK* cells, cell fluorescence and *sstorA–gfp* mRNA levels returned to near wild-type (wt) levels (Fig. S1A and B).

To identify the factor(s) responsible for disappearance of *sstorA–gfp* mRNA, we screened a mini-Tn*10* transposon library in the Δ*dnaK* strain background to isolate suppressor mutants that restored ssTorA–GFP expression, and thus cell fluorescence. The isolated suppressors mapped to the *cas* operon and, in particular, to the endoribonuclease encoded by *casE* and the promoter region of the HD nuclease-helicase encoded by *cas3* (Fig. 5.1.1). A 'clean' deletion of the *casE* gene in strain BW25113 Δ*dnaK* was created. As expected, cell fluorescence was recovered in Δ*dnaK*Δ*casE* cells expressing ssTorA–GFP (Fig. 5.1.1 and C). Fluorescence in this strain was attributed to stable accumulation of *sstorA–gfp* mRNA (Fig. 5.1.1). To ensure this phenotype was due to *casE* inactivation and not polar effects on surrounding genes, we

complemented the Δ*dnaK*Δ*casE* strain with a plasmid-encoded copy of CasE. Following expression of ssTorA–GFP, Δ*dnaK*Δ*casE* cells that had been genetically complemented with CasE were only weakly fluorescent (Fig. S1A) and accumulated a correspondingly low level of *sstorA–gfp* mRNA (Fig. S1B), indicating that CasE was responsible for the gene-silencing phenotype. Expression of an H20A variant of CasE ($CasE^{H20A}$) that is incapable of pre-crRNA cleavage (Brouns et al., 2008) did not confer ssTorA–GFP silencing (Fig. S1A). The *E. coli* K12 *cas* system encodes at least eight genes: *cas1*, *cas2*, *cas3* and five genes designated *casABCDE* that encode the Cascade complex (Brouns et al., 2008) (Fig. 5.1.1). Individual deletion of each of these genes in a Δ*dnaK* strain background revealed that all of the other *cas* pathway components were required for the gene-silencing phenotype (Fig. S2A). It should be noted that the absence of DnaK was obligatory for silencing because wt and single *cas* mutants in the wt background all emitted strong fluorescence following ssTorA–GFP expression (data not shown). The requirement of the Cas enzymes for silencing was corroborated by stable accumulation of ssTorA–GFP in BL21(DE3) Δ*dnaK* cells (Fig. S2B), which lack the Cas enzymes (Brouns et al., 2008) but encode two CRISPR loci (Grissa et al., 2007) with several spacers that are complementary to *sstorA* (Fig. S2C).

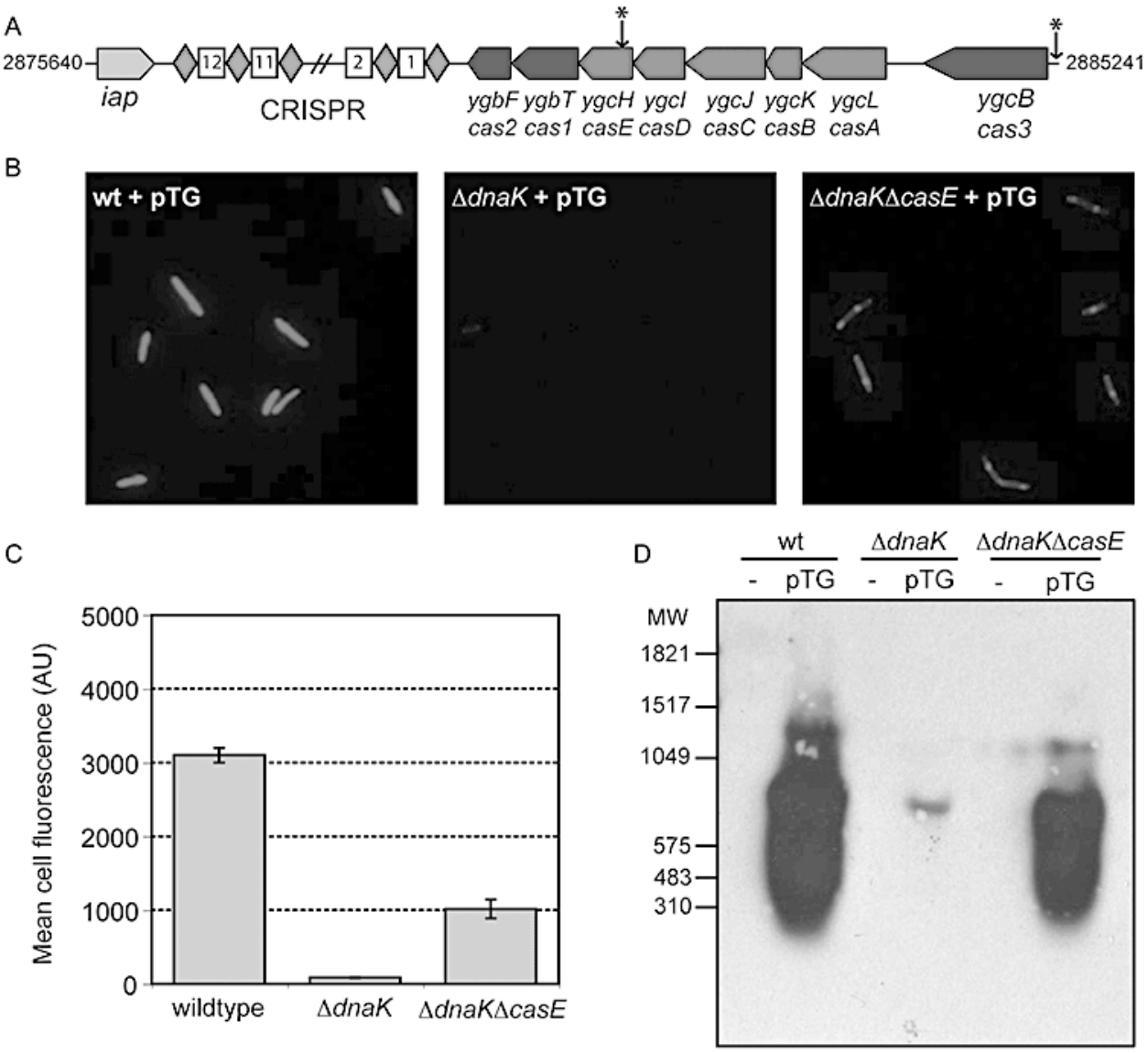

Figure 5.1.1 CasE-dependent silencing of ssTorA–GFP. (a) The CRISPR-Cas pathway of *E. coli* K12. Repeats and spacers are indicated by diamonds and rectangles respectively. Protein family nomenclature is described in elsewhere (Haft et al., 2005). Asterisk indicates the relative location of the mini-Tn*10* insertion sites in *casE* and upstream of *cas3* that were isolated by transposon library screening. B and C. (b) Fluorescence microscopy and (c) flow cytometric screening of *E. coli* BW25113 (wt), BW25113 Δ*dnaK* or BW25113 Δ*dnaK*Δ*casE* expressing ssTorA–GFP from pTG, as indicated. D. Northern blot of total RNA isolated from BW25113 (wt), BW25113 Δ*dnaK* or BW25113 Δ*dnaK*Δ*casE* cells carrying no plasmid (–) or expressing ssTorA–GFP from pTG. Blot was probed using GFP-specific digoxigenin-labelled oligonucleotides. Experiments performed in this figure by R. Perez-Rodriguez.

We next investigated whether ssTorA–GFP silencing was dependent on complementarity between genomic spacers and plasmid-encoded *sstorA*. The *E. coli* K12 genome contains two CRISPR loci, the first of which is adjacent to the *cas* operon and includes three spacers (spacers 1, 5 and 8) that contain sequence complementarity to different regions of *sstorA* (Fig. 5.1.2A). It is noteworthy that all of these spacers were present in the genome prior to the introduction of plasmid pTG; thorough sequencing of both CRISPR regions prior to and at many time points (ranging from 1 h to as long as 3 days) after the introduction of the pTG plasmid provided no evidence for the insertion of new spacers corresponding to pTG. Nonetheless, when the entire CRISPR locus adjacent to the Cas enzymes (Fig. 5.1.1A) was deleted in Δ*dnaK* cells, silencing of *sstorA–gfp* was abolished as evidenced by the restoration of fluorescence to a level that was nearly identical to that observed in the Δ*dnaK*Δ*casE* mutant (data not shown). This suggested that silencing in Δ*dnaK* cells was indeed CRISPR-dependent. To determine whether complementarity between *sstorA* and spacers 1, 5 and/or 8 was responsible for silencing, we modified plasmid-encoded *sstorA* with synonymous mutations that we predicted would reduce base pairing between the crRNA and complementary target sequence, but would not change the ssTorA sequence at the protein level. Two or six synonymous point mutations were introduced into *sstorA* yielding the constructs ssS5–GFP or ssS1.5.8–GFP respectively (Fig. 5.1.2A). The synonymous mutations did not affect expression or localization of these constructs in wt cells relative to unmodified ssTorA–GFP (Fig. 5.1.2B). However, following expression in Δ*dnaK* cells, only ssTorA–GFP but not ssS5–GFP or ssS1.5.8–GFP was silenced (Fig. 5.1.2B and D), indicating that as few as two base pair mismatches was sufficient to block silencing. These results indicate that ssTorA–GFP silencing is specified by sequence identity between the

encoding plasmid and a genomic spacer. It should be noted that despite the relatively short sequence matches, the complementary sequences were all GC rich (Fig. 4.1.2A).

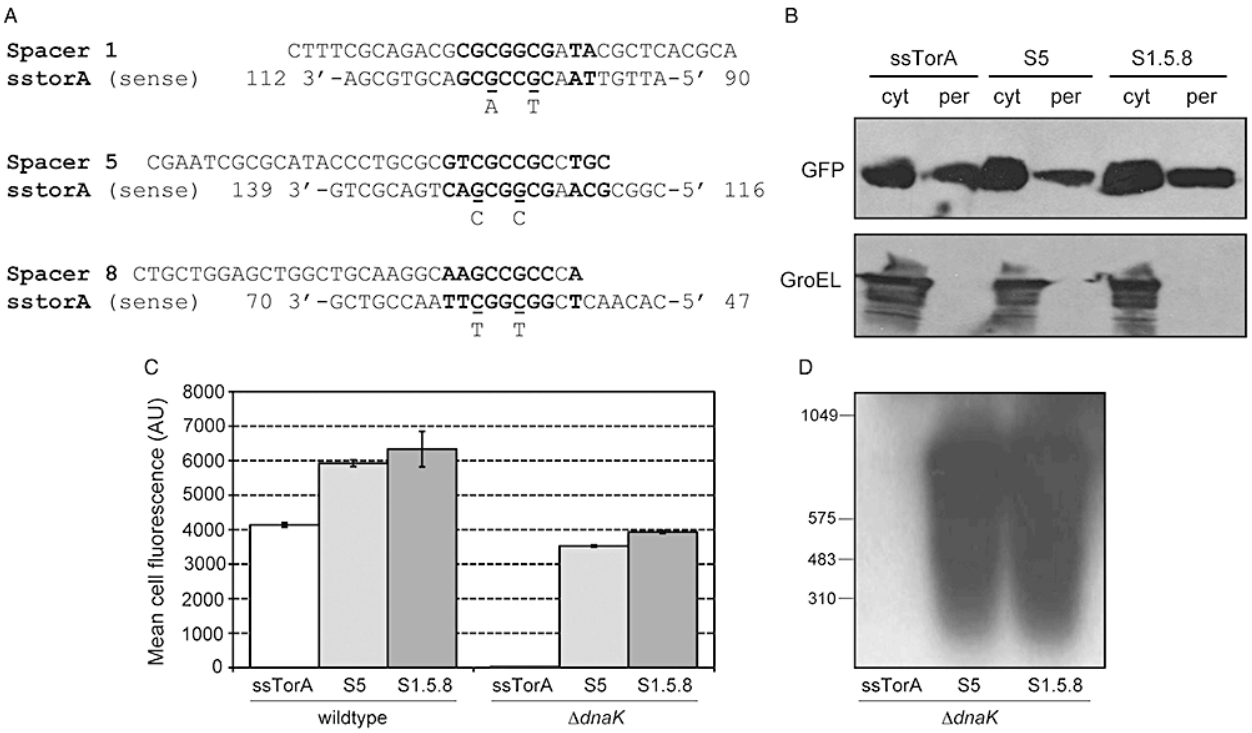

Figure 5.1.2 Complementarity between spacer DNA and ssTorA sequence is required for silencing activity. A. *Escherichia coli* K12 spacers 1, 5 and 8 aligned with the sense strand of *sstorA*. Bold letters correspond to complementary base pairs. Numbering of *sstorA* corresponds to the bp position relative to the start codon of native *E. coli* TorA. Synonymous base changes are indicated below the *sstorA* sequence. The ssS5–GFP construct was created by mutating the C57 and C60 nucleotides in *sstorA* to T57 and T60 respectively. The ss1.5.8–GFP construct was created by further mutating ssS5–GFP as follows: G99T, G102A, G126C and G129C. B. Western blot of the cytoplasmic (cyt) and periplasmic (per) fractions generated from BW25113 wt cells expressing ssTorA–GFP, ssS5–GFP or ssS1.5.8–GFP. Blot was probed with anti-GFP antibodies, stripped and reprobed with anti-GroEL antibodies, where GroEL served as a fractionation marker. C. Flow cytometric screening of BW25113 wt and Δ*dnaK* cells expressing ssTorA–GFP, ssS5–GFP or ssS1.5.8–GFP. D. Northern blot of total RNA from BW25113 Δ*dnaK* cells expressing ssTorA–GFP, ssS5–GFP or ssS1.5.8–GFP. Blot was probed using GFP-specific digoxigenin-labelled oligonucleotides. Experiments performed in this figure by R. Perez-Rodriguez.

5.2 Reconstitution of the core Cas complex

As all of the *cas* genes tested were required for silencing (Fig. S2A) and as the CasABCDE proteins were previously shown to copurify (Brouns et al., 2008), we next attempted to reconstitute the minimal *E. coli* Cas complex needed to process the pre-crRNAs into crRNAs. Each Cas protein was purified individually and most behaved as monomers in solution including CasE, which migrated at ~30 kDa on the sizing column and interacted strongly with nucleic acids based on strong absorption at UV_{260} (Fig. 5.2A). CasC, on the other hand, oligomerized into a ~490 kDa species. Interestingly, CasD was only soluble when coexpressed with CasC and the resulting complex migrated as a 105 kDa species (Fig. 5.2A) with an estimated molar ratio of 2:1 for CasC and CasD, respectively, based on Coomassie-stained band intensities (Fig. 5.3B). Strong interactions were also detected between CasC and CasE, which formed a ~1:1 molar ratio complex migrating at ~832 kDa (Fig. 5.2A and C). Different reconstitution schemes confirmed that CasC, CasD and CasE further formed a salt-stable ternary complex that was ~200 kDa when CasD was fused with an N-terminal SUMO tag (Fig. 5.2A). Solubility was reduced in the absence of the SUMO tag, and the ternary complex migrated as a > 600 kDa species with an apparent stoichiometry estimated at 6:1:1 for CasC, CasD and CasE respectively (Fig. 5.2D). CasC appears to be the centre of the complex as no interactions were found between CasD and CasE. Previously, CasA and CasB were found at substoichiometric levels in the affinity purified *E. coli* Cascade complex (Brouns et al., 2008); however, incubation of separatedly purified CasA with the CasCDE complex did not lead to higher complex formation under 'physiological' conditions. Only marginal interactions were observed between CasB and CasC under low-salt conditions, which diminished under stringent conditions. Thus, we conclude that CasA and CasB are not stable components of the core Cas complex. This, however, does not rule out the

possibility that CasA and CasB associate with the core Cas complex in a nucleic acid-dependent fashion. To investigate whether the molecular composition of the CasCDE complex is conserved phylogenetically, we carried out similar reconstitution of *Thermus thermophilus* Cas proteins and observed an identical ternary complex formed by CasCDE (data not shown). Furthermore, *T. thermophilus* CasE, which is ~35% identical to *E. coli* CasE, was found to hetero-assemble with *E. coli* CasC or CasCD to form salt-stable higher complexes (data not shown). Thus, CasCDE-containing bacteria appear to form a conserved Cas core complex in Cas system subset 2 (Brouns et al., 2008).

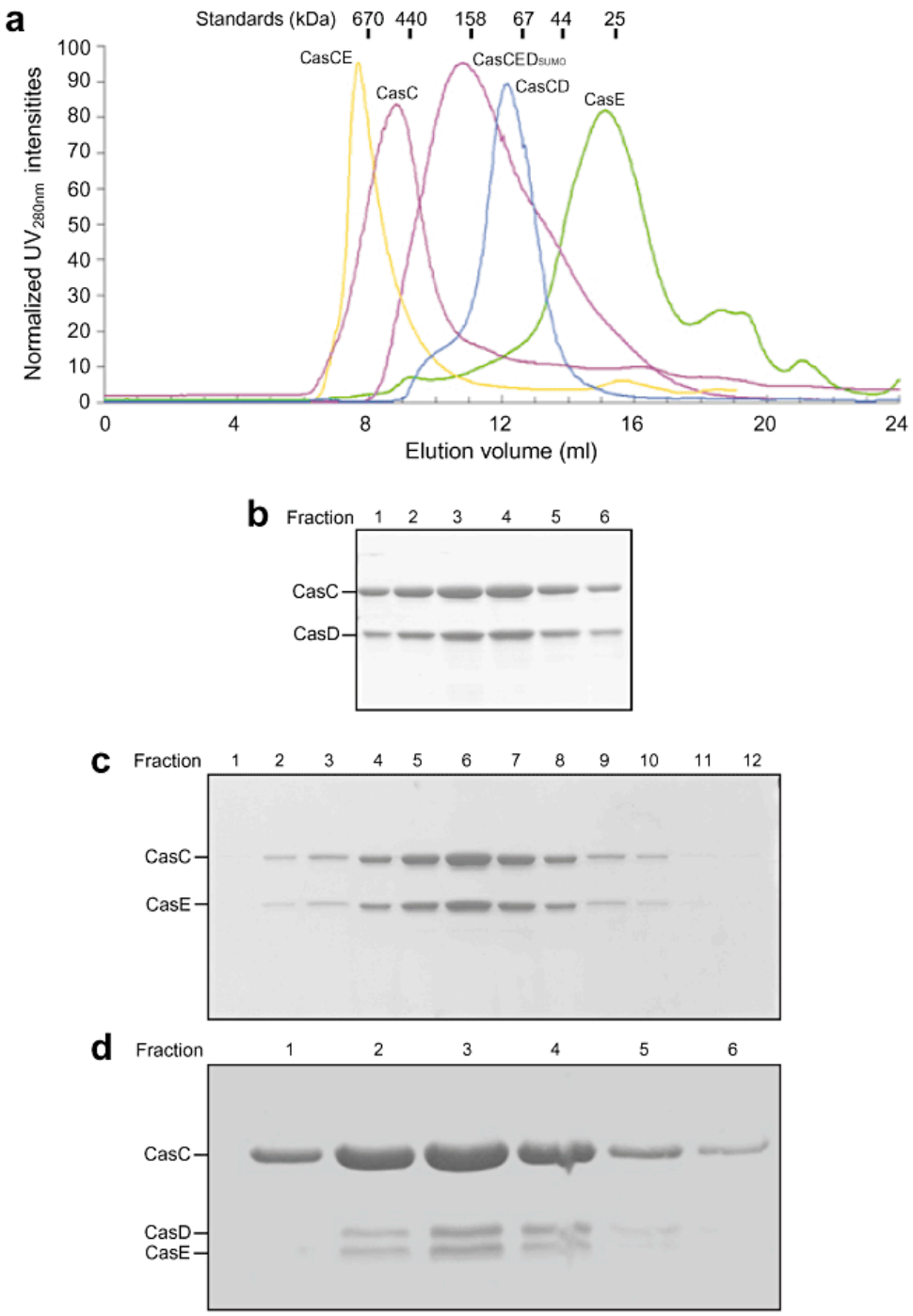

Figure 5.2 Reconstitution of *E. coli* Cascade core complexes. A. Elution profile of CasC and CasE proteins and CasCD, CasCE, and CasCDE complexes on analytical Superdex 200 size-exclusion column. The *x*- and *y*-axes depict elution volume and UV absorbance at 280 nm respectively. Nucleic acids have been removed from Cas protein/complexes on prior ion-exchange columns. B–D. SDS-PAGE analysis of the size-exclusion fractions of the (B) CasCD, (C) CasCE and (D) CasCDE complexes. Experiments performed in this figure by Q. Huang and K. Nam.

5.3 CRISPR RNA processing by core Cas complex

To reveal the function of the individual Cas proteins or Cas complexes in CRISPR RNA processing, we carried out *in vitro* RNA processing assays. Previous studies showed that pre-crRNA processing into unit length in *E. coli* requires a subset of Cas proteins, especially a catalytically competent CasE (Brouns et al., 2008; Carte et al., 2008). CasE-mediated processing involves an invariable cleavage site at the base of the 5′ hairpin structure to release the mature crRNA and a variable cleavage site at three to four different positions in the vicinity of the hairpin loop that may be due to the non-specific action of cellular ribonucleases (Brouns et al., 2008; Carte et al., 2008). Here, a HEX-labelled CRISPR RNA containing the conserved CRISPR repeat was used as substrate. When individually purified Cas proteins were incubated with this 5′-HEX-labelled CRISPR RNA, only CasE showed strong RNase activity (data not shown). By comparison with the alkaline hydrolysis ladder on sequencing gels, we determined that CasE cleaved at the base of the CRISPR stemloop (Fig. 5.3), coinciding with the invariable cleavage site that defines the 5′ end of the processed crRNA (Brouns et al., 2008). This activity is specific for CasE as it disappeared when the CRISPR RNA was incubated with the $CasE^{H20A}$ catalytic mutant (Fig. 5.4). We then carried out a more thorough analysis of the cleavage products produced by CasCE and the CasCDE complexes (in which CasD is N-terminally tagged with SUMO). Both of these complexes were shown to generate the same cleavage pattern on the CRISPR repeat RNA as the CasE protein alone (Fig. 5.3). The enzymatic activity again came from CasE, as $CasCE^{H20A}$ was unable to cleave the CRISPR repeat RNA. These results suggested that formation of higher order complexes did not alter the enzymatic specificity of CasE protein. We also verified that the cleavage patterns of the SUMO-tagged complex and untagged CasCDE complex, which migrated as a large oligomeric species, were identical in our assay conditions

(data not shown) thus eliminating the possibility that the N-terminal SUMO tag in CasD altered the enzymetic specificity of the CasCDE complex. We cannot, however, rule out the possibility that the oligomeric form of the CasCDE complex is more efficient in processing the multimeric form of the pre-CRISPR RNA because of more efficient substrate binding.

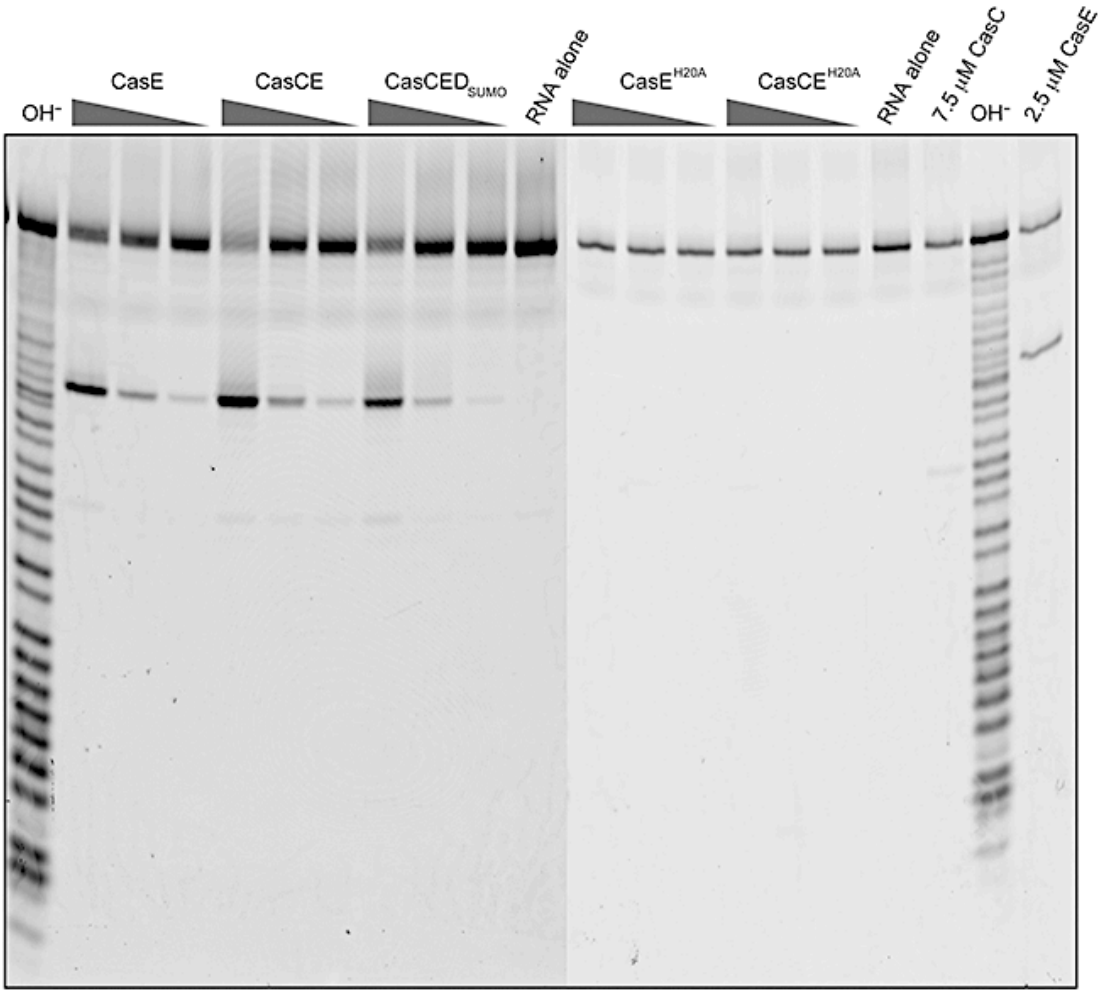

Figure 5.3 RNase activity of Cas proteins and CasE-containing complexes. Cleavage profile of CasE, CasCE, CasCED$_{SUMO}$, CasE^{H20A}, CasCE^{H20A} and CasC alone resolved on 18% sequencing gel. The amount of protein used in the assay was from left to right 2.5, 0.5 and 0.1 μM. Alkaline hydrolysis ladder in lanes 1 and 20 resolves the position of every base on the 18% sequencing gel. Experiments performed in this figure by Q. Huang and K. Nam.

5.4 Cas-mediated ssTorA–GFP silencing occurs via targeting of plasmid DNA

We next investigated whether plasmid pTG encoding ssTorA–GFP was targeted by the Cas machinery in Δ*dnaK* cells. Following induction of ssTorA–GFP, we purified plasmid pTG from an equivalent number of wt, Δ*dnaK* and Δ*dnaK*Δ*casE* cells and performed restriction analysis on the isolated plasmids. Plasmid pTG was stably maintained in wt and Δ*dnaK*Δ*casE* cells but was hardly detectable in Δ*dnaK* cells (Fig. 5.4A). Silencing of pTG in Δ*dnaK* cells was dependent on the *sstorA–gfp* sequence because empty pTrc99A plasmid accumulated at a similar level in all strains including Δ*dnaK* cells (Fig. 5.4A). We reasoned that plasmid silencing would cause a measurable decrease in antibiotic resistance conferred by the *ampR* gene encoded elsewhere on pTG. Indeed, the minimum inhibitory concentration (MIC) on ampicillin (Amp) of wt and Δ*dnaK*Δ*casE* cells was eightfold and sixfold greater than that for Δ*dnaK* cells (Fig. 5.4B). To verify that the Cas machinery only targets plasmid DNA and not 'self' DNA, we determined the viability of wt, Δ*dnaK* and Δ*dnaK*Δ*casE* cells by plating on non-selective medium following ssTorA–GFP induction. The colony-forming units (cfu) of Δ*dnaK* cells expressing ssTorA–GFP was actually slightly greater than the cfu of wt and Δ*dnaK*Δ*casE* cells (Fig. 5.4C), indicating that *E. coli* Cas effector complexes specifically target extrachromosomal DNA while avoiding genomic DNA. To test whether the Cas complex itself is capable of plasmid degradation, we measured the DNase activity of CasE, CasCE and $CasCED_{SUMO}$ against single- and double-stranded DNA oligos, and against supercoiled plasmids including pTG encoding ssTorA–GFP and an unrelated pUC19 plasmid. Under our experimental conditions, neither CasE, nor the CasE-containing complexes had any measurable DNase activity (data not shown).

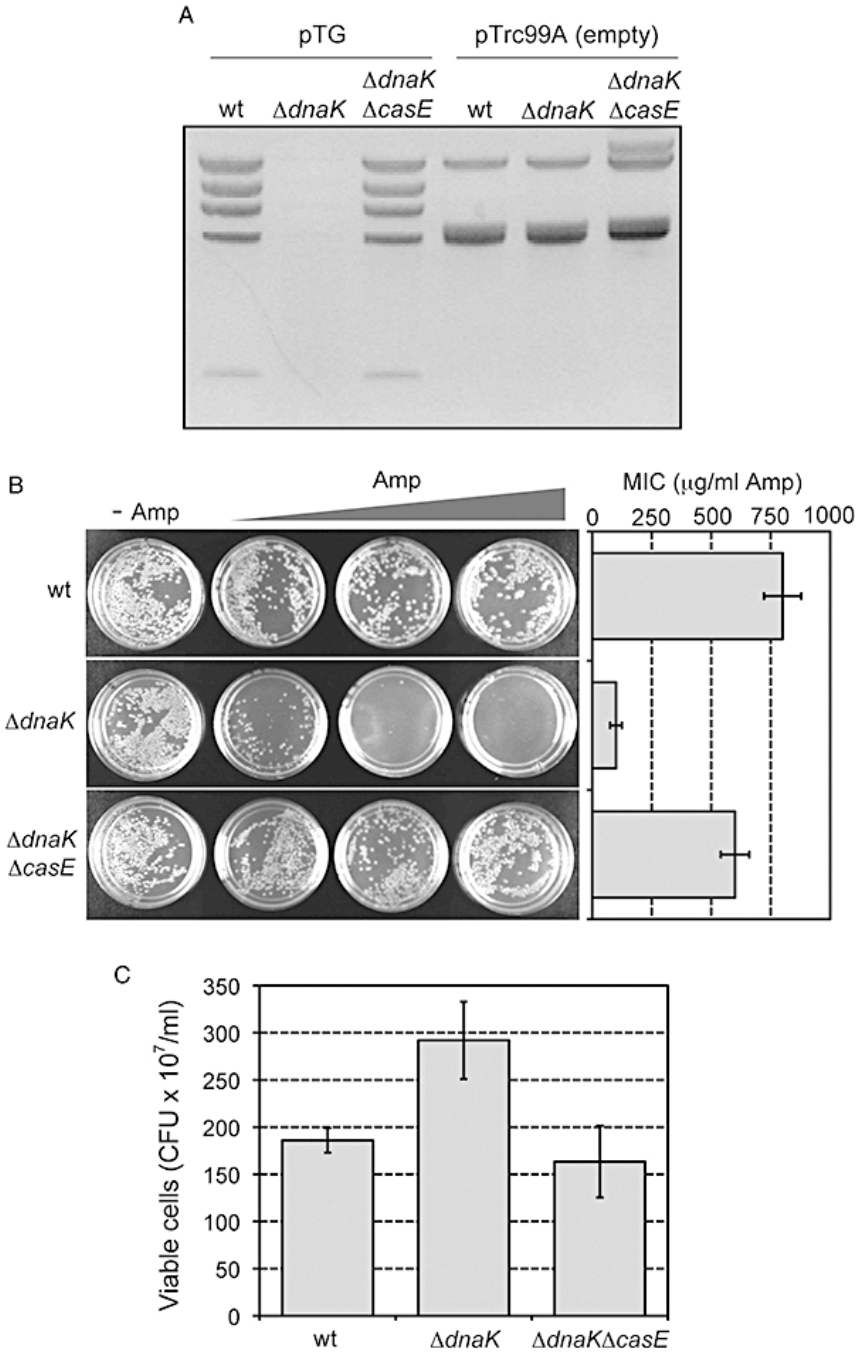

Figure 5.4 Plasmid DNA is the target of CRISPR-Cas interference. A. RsaI restriction analysis of plasmid DNA isolated from BW25113 wt, Δ*dnaK* and Δ*dnaK*Δ*casE* cells transformed with either pTG expressing ssTorA–GFP or empty pTrc99A (control) as indicated. B. Selective plating of BW25113 wt, Δ*dnaK* and Δ*dnaK*Δ*casE* cells expressing ssTorA–GFP. An equivalent number of cells was plated on LB agar with no Amp (– Amp) or supplemented with 50, 100 and 200 μg ml^{-1} Amp. C. Viability of the same cells as in (B). The number of cfu was determined by plating an equivalent number of induced cells on LB agar with no antibiotics. Experiments in this figure performed by R. Perez-Rodriguez.

5.5 Cas-mediated silencing depends on the BaeSR two-component signal transduction system

We next sought to determine how the CRISPR-Cas system was activated in Δ*dnaK* cells. As GFP expressed without a Tat export signal is not silenced in cells that lack DnaK (Pérez-Rodríguez et al., 2007), targeting of GFP to the Tat translocase is a requisite for Cas-mediated silencing in Δ*dnaK* cells. Indeed, when the twin arginines in the Tat signal peptide of ssTorA–GFP were replaced with twin lysines [TorA(R11K:R12K)], amino acid substitutions that abolish Tat-dependent GFP export (DeLisa et al., 2002), no silencing of ssTorA–GFP was observed in Δ*dnaK* cells (Fig. 5.5A and B). Likewise, there was also no measurable silencing when export-competent ssTorA–GFP was expressed in Δ*dnaK* cells that also lacked the TatC translocase component, which is required for export (Fig. 5.5A and B).

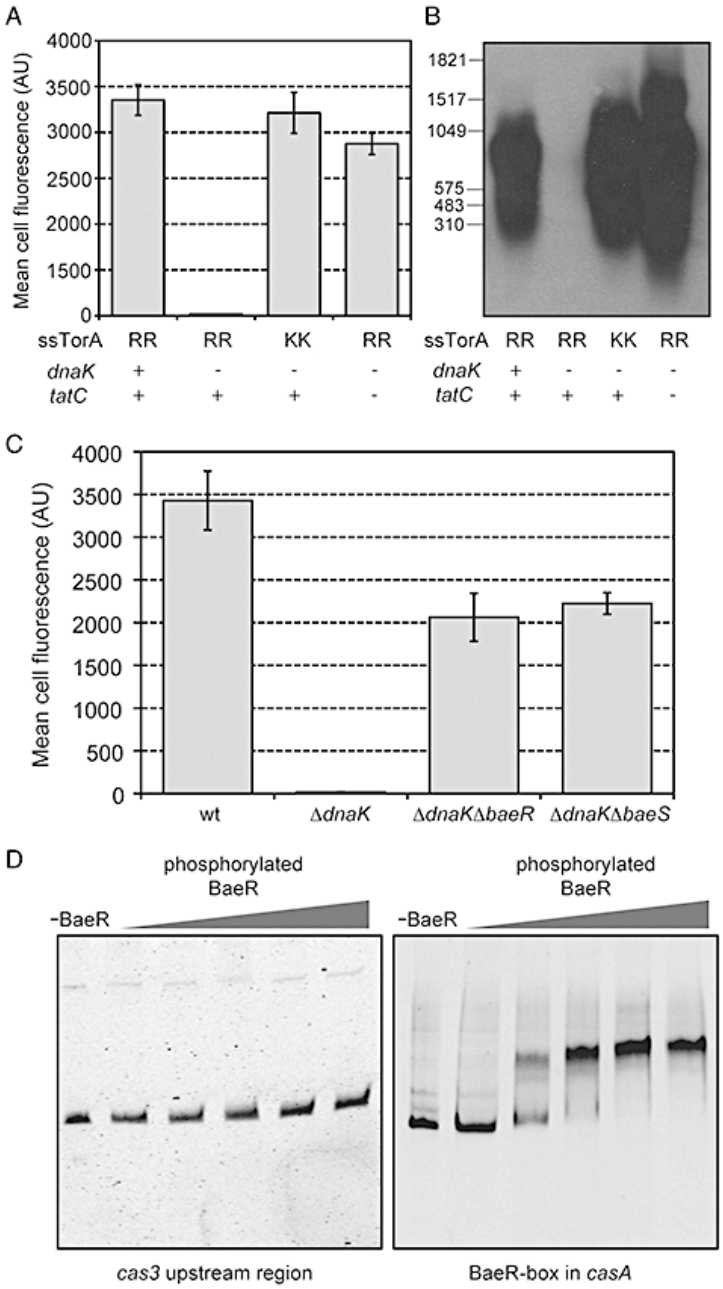

Figure 5.5 CRISPR-Cas silencing is dependent on membrane targeting and the BaeSR two-component stress response system. A. Flow cytometric screening of BW25113 wt, Δ*dnaK* and Δ*dnaK*Δ*tatC* cells expressing either ssTorA–GFP or ssTorA(KK)–GFP as indicated. B. Northern blot analysis of total RNA from the same cells as in (A). Blot was probed using GFP-specific digoxigenin-labelled oligonucleotides. C. Flow cytometric screening of BW25113 wt, Δ*dnaK*, Δ*dnaK*Δ*baeS* and Δ*dnaK*Δ*baeR* cells expressing ssTorA–GFP. D. Electrophoretic mobility shift assay for BaeR binding to the Cas promoter region. A total of 0.15 pmol of DNA fragments corresponding to *cas3* (415 bp) and *casA* (1057 bp) promoter regions was incubated without (–BaeR) or with various concentrations of phosphorylated BaeR-His_6. The amount of BaeR was from left to right 9.2, 11.5, 16.1, 24.2 and 32.25 μM. Experiments performed in this figure by R. Perez-Rodriguez and C. Haitjema.

Based on these findings, we hypothesized that membrane localization of the highly expressed ssTorA–GFP fusion in Δ*dnaK* cells might induce an envelope stress, perhaps because of mislocalization of GFP in the inner membrane in these cells. We further speculated that signal transduction pathways that control the adaptive response to envelope stresses in *E. coli* might activate the expression of *cas* genes. If this interpretation is correct, then we would predict little to no *cas* gene expression, and thus no silencing activity, following ssTorA–GFP expression in cells where envelope stress response systems are compromised. To test this, we investigated the BaeSR two-component regulatory system, which includes a membrane-localized histidine kinase BaeS that senses envelope stress and modulates the phosphorylation state of the cytoplasmic transcription factor BaeR (MacRitchie et al., 2008). We focused our attention on this signalling pathway because overexpression of BaeR was previously shown to activate the promoter of the *casA* gene (Baranova and Nikaido, 2002). In support of our hypothesis above, we observed stable accumulation of ssTorA–GFP fluorescence in Δ*dnaK* cells that also lacked BaeS or BaeR (5.5C).

As BaeSR was genetically involved in transmitting the envelope stress caused by ssTorA–GFP expression, we next investigated whether the BaeR regulator was able to bind *cas* promoter DNA. Recently, DNase I footprinting studies identified the BaeR-box sequence TCTNCANAA, where N is any nucleotide (Yamamoto et al., 2008). Our own inspection of the *cas* operon revealed a putative BaeR-box sequence of TCTGCATAA within the coding region of *casA* (nucleotides 298–306) and upstream of *casBCDE*. To determine whether BaeR was capable of binding this sequence, we performed electrophoretic mobility shift assays using purified BaeR and a 1057 bp DNA fragment covering the BaeR-box sequence. Two additional 358 and 415 bp

DNA fragments upstream of *casA* and *cas3* were also tested as candidate BaeR binding regions. As a positive control, a 279 bp fragment including the *acrD* promoter, a known binding target of BaeR (Hirakawa et al., 2005), was used. BaeR-His_6 protein was purified and phosphorylated with acetylphosphate. The electrophoretic mobility of the DNA fragments corresponding to the *acrD* promoter region (data not shown) and the BaeR-box sequence internal to *casA* (Fig. 5.5D) shifted upon the addition of phosphorylated, but not unphosphorylated, BaeR. In contrast, no interaction was observed between phosphorylated BaeR-His_6 and the DNA fragments derived from regions upstream of *casA* (data not shown) or *cas3* (5.5D). Thus, complexes formed by BaeR-His_6 and the BaeR-box sequence internal to *casA* detected in this experiment were specific.

5.6 Induction of cas genes in ΔdnaK cells expressing ssTorA–GFP

We next determined how different mutations in the chromosome affected *cas* gene expression when ssTorA–GFP was induced. As expected, upregulation of *casD* and *casE* was observed following ssTorAGFP expression in Δ*dnaK* cells but not wt cells as determined by quantitative reverse transcription polymerase chain reaction (qRT-PCR; Fig. 5.6). It should be noted that in *dnaK* mutants, *casD* and *casE* expression was induced 15 min after ssTorA–GFP expression (Fig. 5.6) but returned to uninduced levels by 2 h after ssTorA–GFP expression (data not shown), indicating that *cas* gene expression is turned off once the envelope stress caused by pTG is eliminated by CRISPR-Cas silencing. However, similar stress adaptation was not observed in a Δ*dnaK*Δ*casE* double mutant. In these cells, *casD* expression remained elevated as long as 2 h after induction of ssTorA–GFP (data not shown), suggesting that the lack of pTG silencing in these cells results in a sustained stress and a corresponding sustained induction of the Cas pathway. Importantly, the induction of both *casD* and *casE* was abolished in Δ*dnaK*Δ*baeR*

or Δ*dnaK*Δ*baeS* double mutants (Fig. 5.6), indicating that the BaeSR pathway is required for induction of the *cas* genes in response to ssTorA–GFP expression in Δ*dnaK* cells. This result is entirely consistent with our observation above that BaeS and BaeR are each required for plasmid silencing. As upregulation of *cas* gene expression depended on BaeSR, we speculated that genes whose induction is mediated by the BaeSR pathway, such as *spy* (Raffa and Raivio, 2002), might also be upregulated in Δ*dnaK* cells expressing ssTorA–GFP. In line with this hypothesis, we observed that *spy* expression was induced in Δ*dnaK* cells expressing ssTorA–GFP (Fig. 5.6). As was seen for *casD* and *casE*, mutation of *baeR* or *baeS* eliminated the elevated *spy* expression that was seen in the Δ*dnaK* strain expressing ssTorA–GFP (Fig. 4.6). However, *spy* expression was diminished in Δ*dnaK*Δ*casE* cells, indicating that the stress caused by expression of ssTorA–GFP in cells where silencing is absent was not a trigger of the BaeSR response. Hence, we conclude that the requirement for the Bae response in cells carrying the induced *sstorA–gfp* and a Δ*dnaK* mutation is complex and involves additional factors, beyond a simple *sstorA–gfp*-mediated upregulation of the Bae response.

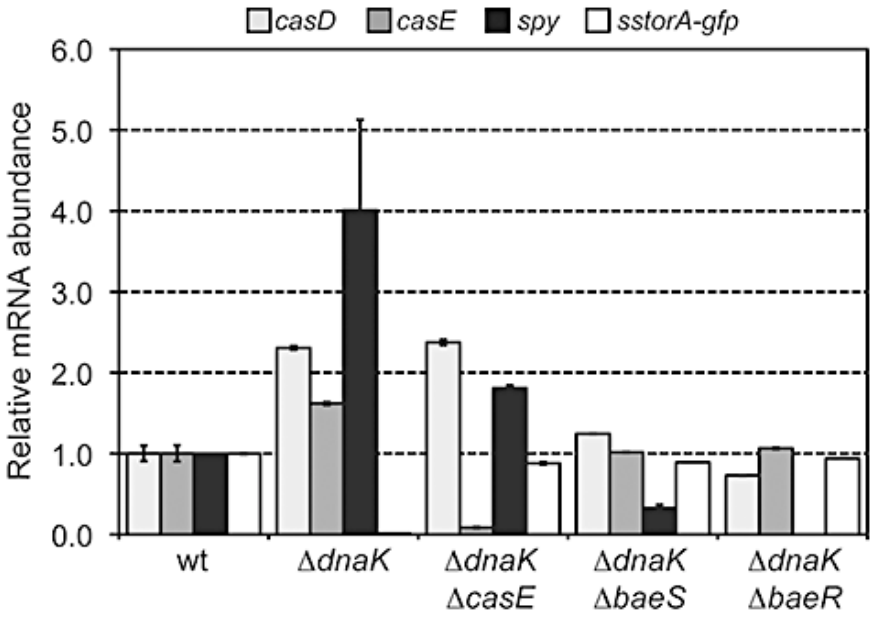

Figure 5.6 Induction of *cas* pathway in response to ssTorA–GFP envelope stress. Expression of *casD* (light gray bars), *casE* (dark gray bars), *spy* (black bars) and *sstorA–gfp* (white bars) as determined by qRT-PCR analysis. Total RNA was isolated 15 min after induction of ssTorA–GFP from each of the strains indicated. All data were normalized to the amount of 16S rRNA measured in each strain. Relative induction was then calculated by the comparative C_t method using the values measured in wt cells expressing ssTorA–GFP as calibrators. It should be noted that the expression levels of *sstorA–gfp* measured in these qRT-PCR experiments were entirely consistent with our Northern blot and cell fluorescence analysis. Experiments performed in this figure by C. Haitjema.

We next examined whether other known Bae inducing cues triggered upregulation of the Cas pathway. Specifically, we tested whether induction of *casD* and *CasE* expression was observed in response to spheroplasting and separately, to the formation of misfolded pilin subunits caused by PapG overexpression. Both of these stresses are well known to activate the Bae signal transduction pathway (Raffa and Raivio, 2002). However, we found no evidence for *casD* or *casE* induction in response to either of these envelope stresses under the conditions tested here (data not shown).

5.7 Functional reconstitution of ssTorA–GFP silencing in cas-deficient E. coli

Finally, we speculated that overexpression of the Cas enzymes from an inducible promoter would bypass the BaeSR-dependent *cas* regulation and produce silencing in a manner that did not require DnaK. BL21(DE3) cells were used to test this notion because they lack genomic copies of the *cas* genes but contain CRISPR loci that encode several spacers with complementarity to *sstorA*. Indeed, coexpression of the entire *cas* operon from pWUR399 (Brouns et al., 2008) along with ssTorA–GFP in BL21(DE3) led to elimination of ssTorA–GFP in both the presence and absence of DnaK (Fig. S2B). This silencing was abolished when an H20A substitution was introduced to *casE* in pWUR399 (data not shown). Consistent with their role as a minimal RNA processing unit, coexpression of the CasCDE enzymes from pWUR402 was insufficient to produce silencing in this system (Fig. S2B). This result indicates that additional enzymes, perhaps from the set of CasA, CasB, Cas1, Cas2 and Cas3, are required for silencing and is in line with our single-gene knockout studies in K12 *E. coli* (Fig. S2A). The ability to functionally reconstitute ssTorA–GFP silencing in BL21(DE3) cells by simple addition of cas genes should enable further dissection of this unique mechanism.

5.8 Cas5d, together with Csd1 and Csd2, complements Cas-deficient E. coli to silence plasmid DNA in vivo.

Like *E. coli,* it was recently identified that *Bacillus halodurans* C-125 also encodes a CRISPR-Cas system, albeit of the Dvulg subtype. Among the Cas enzymes in the *B. halodurans* CRISPR-Cas system is the Cas5d enzyme, which, like *E. coli* Cas5e (CasD), belongs to the Cas5 family of proteins. The *B. halodurans* also encodes Csd1 and Csd2 which are predicted to be the functional analogs of *E. coli* CasA and *E. coli* CasC, respectively. Thus, we asked the question

whether these *B. halodurans* enzymes can complement the activity of their predicted *E. coli* counterparts. To test the *in vivo* activity of Cas5d and the newly identified Dvulg-Cascade complex, we utilized a previously established fluorescence-based genetic reporter of CRISPR-Cas interference in *E. coli* (Perez-Rodriguez et al., 2011). In this earlier study, we observed that GFP targeted to the twin-arginine translocation (Tat) pathway induced the CRISPR-Cas system in *E. coli* cells lacking the molecular chaperone DnaK. Activation of this CRIPSR-Cas response was dependent on the BaeSR two-component signaling pathway, which is typically involved in cell envelope stress responses. As a result, plasmid DNA encoding the Tat-targeted GFP reporter, namely ssTorA-GFP, was silenced and GFP fluorescence was abolished. This silencing phenotype required each Cas protein (i.e., Cascade and Cas1-3) and the genomic CRISPR region adjacent to the Cas operon; deletion of any of these components abolished the plasmid silencing. For example, silencing-competent Δ*dnaK* cells expressing ssTorA-GFP were virtually non-fluorescent but when the gene encoding one of the Cas proteins, such as CasA (Cse1), CasB (Cse2), CasC (Cse4), CasD (Cas5e), or CasE (Cse3) was deleted, silencing activity was abolished and cells became highly fluorescent (Fig. 5.8). Using this genetic system, we attempted to complement the different *E. coli* Cas mutants with the functional or sequence homolog from *B.halodurans*, namely Csd1 for CasA, Csd2 for CasC, and Cas5d for CseD and CasE. Only the *B.halodurans* Csd2 enzyme was able to restore silencing activity to the corresponding CasC mutant *E. coli* strain (Fig. 5.8). Interestingly, simultaneous coexpression of Cas5d, Csd1, and Csd2 resulted in a strong silencing phenotype in all of the mutant strain backgrounds (Fig. 5.8), consistent with our biochemical reconstitution data showing that these proteins form a Cascade like assembly. Based on our study, the *E. coli* pre-crRNA is expected to be a poor substrate for *B. halodurans* Cas5d. However, the residual processed crRNA may be sufficient to render plasmid

silencing together with the overexpressed *B. halodrans* Cascade. Taken together, these results suggest that the Dvulg-Cascade complex formed in vivo is functionally active, using extrachromosomal DNA as the target.

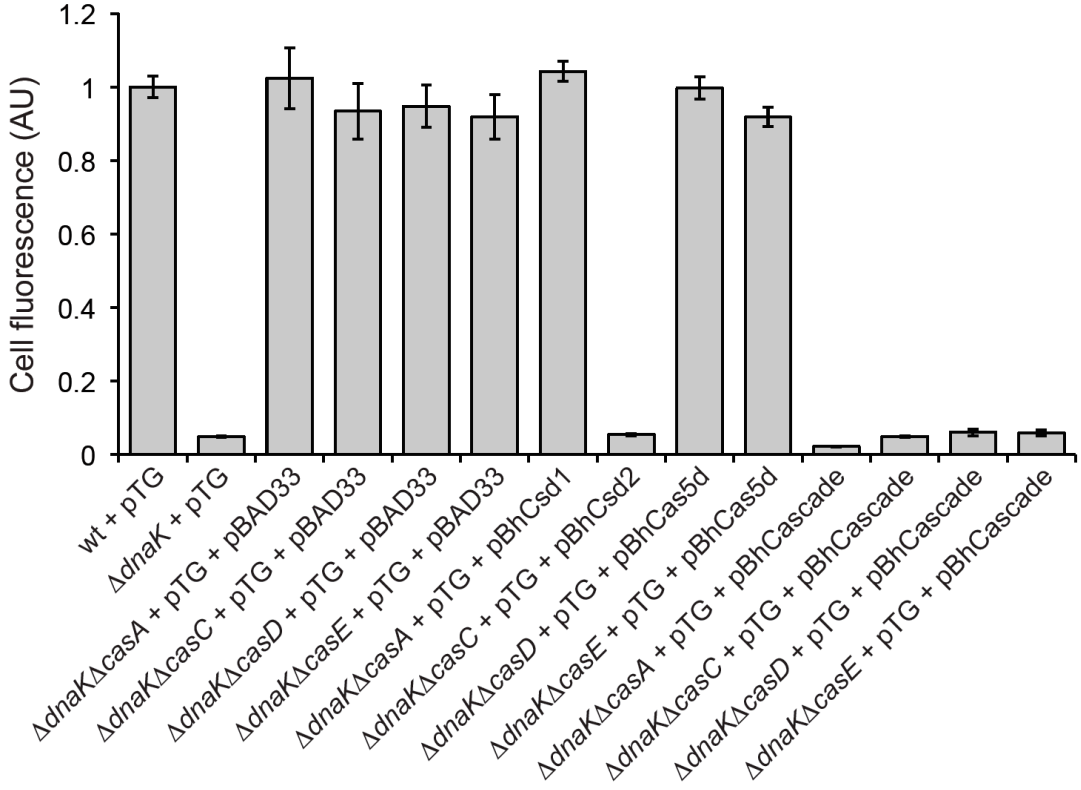

Figure 5.8 Cas-dependent silencing of ssTorA-GFP in *E. coli*. Cellular fluorescence of wild-type (wt) *E. coli* BW25113, BW25113 Δ*dnaK*, and isogenic cas mutant strains (e.g., BW25113 Δ*dnaK*Δ*casE*) expressing ssTorA-GFP from plasmid pTG. Complementation of silencing activity was assayed using pBhCsd1, pBhCsd2, pBhCas5d or pBhCascade (Cas5d/Csd1/Csd2) and compared to the empty vector (pBAD33) control. Data is the average of three replicate experiments and error is reported as the standard error of the mean. Experiments performed in this figure by C. Haitjema.

5.9 Discussion

Previous studies have firmly established that CRISPR-Cas systems protect bacteria by conferring immunity against bacteriophage infection (Barrangou et al., 2007; Brouns et al., 2008) and limiting plasmid conjugation (Marraffini and Sontheimer, 2008). Here, we made the

unexpected discovery that membrane localization of ssTorA–GFP in Δ*dnaK* cells activated the Cas pathway and led to silencing of the ssTorA–GFP-encoding plasmid. At present, the reason for Cas activation following ssTorA–GFP expression in Δ*dnaK* cells is poorly understood. The Tat system exports proteins that have undergone folding in the cytoplasm (Sanders et al., 2001; DeLisa et al., 2003)by inserting these folded substrates into the inner membrane (Hou et al., 2006). However, when aberrantly folded proteins are targeted to the Tat system, cells exhibit severe growth defects that are likely due to proton leakage at the membrane (Richter and Brüser, 2005). Hence, we speculate that the absence of the DnaK chaperone, which is known to assist the folding and translocation of certain Tat substrates (Oresnik et al., 2001; Perez-Rodriguez et al., 2011), resulted in misfolding and/or mislocalization of ssTorA–GFP in a manner that disrupted membrane integrity or was otherwise toxic to cells. The subsequent activation of CRISPR-Cas suggests that this system may provide a defence mechanism against defective protein localization that threatens the integrity of the inner cytoplasmic membrane. It is worth mentioning that we have not been able to find any evidence for the acquisition of new genomic spacers derived from plasmid pTG. Instead, the spacers targeting sstorA were encoded in existing genomic spacers. Whether preexisting spacers result in CRISPR-mediated silencing of endogenous torA or any other endogenous proteins is currently unclear. However, it seems unlikely that torA silencing is a conserved mechanism given the enormous variety in CRISPR spacers (Díez-Villaseñor et al., 2010).

Nonetheless, our results represent the first evidence for the induction of CRISPR-Cas activity by a cue other than phage infection or plasmid conjugation (Barrangou et al., 2007; Brouns et al., 2008). While these cues are different, they are known to elicit overlapping stress responses (MacRitchie et al., 2008). Thus, we speculate that activation of CRISPR-Cas in

response to defective protein export, phage infection and plasmid conjugation may be mediated by the same (or overlapping) stress sensing mechanisms (Fig. 5.9). In this study, ssTorA–GFP silencing in Δ*dnaK* cells was dependent upon the BaeSR signal transduction pathway, indicating that an endogenous cellular mechanism may be involved in the CRISPR-Cas defence system. Such regulated control of Cas enzyme expression is probably needed because constitutive expression of these potent RNases and DNases would be detrimental to the fitness of the host. In fact, recent evidence indicates that cas gene expression is strongly repressed by the nucleoid-associated protein H-NS (Pul et al., 2010), a protein known to silence transcription of laterally acquired genes. H-NS downregulates cas expression by binding to a promoter located upstream of the casA gene and its repression is proposed to be released by anti-silencing mechanisms, which allow the cell to express the genes under specifically regulated conditions. One such anti-silencing mechanism involves the transcriptional activator LeuO (Westra et al., 2010). We show here that another possible anti-silencing mechanism involves the BaeSR two-component regulatory system. Following induction of stress by ssTorA–GFP expression, activated BaeR may promote release of H-NS-mediated cas repression by binding to a newly identified promoter sequence (TCTGCATAA) that is located within the coding region of casA and downstream of the cas promoter where H-NS binds. It should be noted that while activation of the cas pathway minimally involves BaeSR, we cannot rule out the involvement of additional envelope stress response pathways that may work to cooperatively combat envelope stress by triggering CRISPR-Cas. Psp response is a likely candidate because it can be induced in *E. coli* upon filamentous phage infection (specifically phage secretin pIV) and by other membrane-damaging agents (Model et al., 1997), and it has been implicated in mitigating the stress associated with highly expressed Tat substrates (DeLisa et al., 2004).

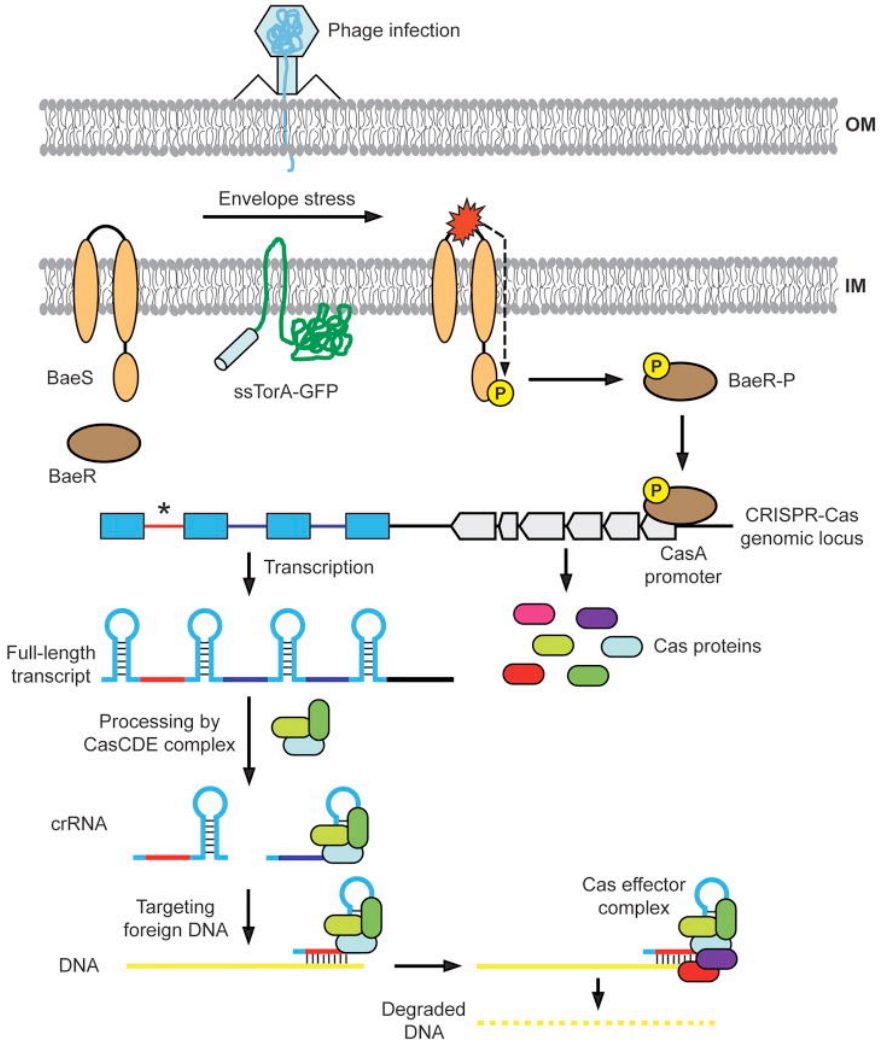

Figure 5.9 Model of CRISPR RNA-mediated DNA silencing in *E. coli*. Following an attack by a phage or the aberrant expression of a secreted or membrane protein, envelope stress response systems, such as the two-component BaeSR signal transduction pathway, are activated. An activated response regulator, for instance phosphorylated BaeR (BaeR-P), triggers expression of the Cas proteins and perhaps also the transcription of full-length CRISPR RNA. Full-length CRISPR RNA is recognized by the CasCDE core complex and processed to produce small guide RNAs (crRNAs), which have a uniformly processed 5′ end and a variable 3′ end. CasE or perhaps the entire CasCDE ternary complex remain associated with the processed crRNAs and may be guided to the target DNA to initiate DNA degradation. Whether CasCDE is a landmark to attract other effector proteins such as CasA and CasB through transient interactions is currently unclear but additional enzymes are necessary as CasE or CasE-containing complexes do not exhibit any measurable DNase activity. Nonetheless, silencing by the Cas effector complex is achieved by base pairing between the crRNA and the DNA target, which may occur during active replication of phage genome or plasmid DNA. Finally, a small number of bacteria acquire new spacers derived from the foreign DNA (marked by an asterisk), leading to survival by CRISPR RNA-mediated degradation of DNA.

Once activated, the CasABCDE proteins reportedly form a complex called Cascade, where the CasE protein within the complex is responsible for processing full-length CRISPR RNA (Brouns et al., 2008). Here we confirmed that CasE is sufficient for pre-CRISPR RNA processing, and that CasE is involved in the formation of a ternary CasCDE core complex that further oligomerizes to form a larger molecular assembly. A mechanistic model can therefore be generated to summarize our data, in which we hypothesize that by assuming a defined tertiary structure, the oligomeric CasCDE core complex possesses greater affinity for multimeric CRISPR repeat regions than individual Cas proteins. This in turn allows more efficient pre-CRISPR processing by CasE. Structure determination of the CasCDE core architecture and its interactions with the CRISPR repeat sequence will be critical for further understanding the pre-CRISPR processing mechanisms.

In addition to *E. coli* K12, we have also observed a CasCDE core in *B. halodurans* and *T. thermophilus* (data not shown), suggesting that the pre-CRISPR processing complex is minimumly conserved in Cas System 2 (CASS2). As CasCD pairs have been identified in five out of the seven categorized CASS subtypes (Makarova et al., 2006), we speculate that these two proteins, especially CasC, likely play an important role in organizing the formation of the Cascade complex. However, it should be noted that the other three Cascade proteins, CasA, CasB and CasE, are not well conserved among CASS subtypes (Makarova et al., 2006). CRISPR repeat sequences and their predicted secondary structures can also differ significantly between organisms (Grissa et al., 2008). Therefore, the pre-CRISPR processing factor in CasE-deficient organisms may differ from that in *E. coli*. Cas6 is found in many CasE-deficient organisms (Haft et al., 2005). Despite sharing little sequence homology, CasE and Cas6 adopt a common

duplicated ferredoxin fold and catalyse metal-independent RNA cleavage (Carte et al., 2008). It remains to be determined whether Cas6 can interact with CasCD in these organisms. Moreover, some bacteria encode neither CasE nor Cas6 in their CRISPR-Cas system. For example, single ferritin fold containing Csy4 protein from Pseudomonas aeruginosa was shown to carry out pre-CRISPR processing in a sequence specific fashion (Haurwitz et al., 2010). Therefore, the pre-CRISPR processing mechanism likely varies among CASS subtypes, underlining the diversity of CRISPR-Cas systems.

While it is now clear that the induction of *cas* gene expression results in the formation of CasCDE, which promotes the maturation of the crRNAs, the recognition and degradation of the foreign target occurs by mechanisms that remain poorly understood. Once generated, crRNAs use base pairing to presumably guide a crRNP complex to the invasive target. In our studies, disruption of perfect base pairing by as few as two mutations was sufficient to abrogate silencing. This is consistent with earlier studies showing that even a single spacer/target mismatch compromises CRISPR interference (Barrangou et al., 2007; Deveau et al., 2008; Marraffini and Sontheimer, 2008). For instance, of 19 phages that evolved to evade CRISPR targeting in *Streptococcus thermophilus*, eight contained a single mutation and three contained a double mutation (Deveau et al., 2008). A surprising aspect of our studies was the relatively short (~8–11 base pairs) spacer/target complementarity that was required for silencing. Given the short lengths of these matches, equilibrium hybridization thermodynamics would seem to be insufficient for discrimination between plasmid DNA and random genomic targets, suggesting that an additional layer of target 'sensing' is performed by the CRISPR-Cas machinery. One example of such an additional layer is the observation that differential complementarity outside of the spacer sequence is an in-built feature of CRISPR systems for self (spacer DNA within the encoding

CRISPR locus itself) versus non-self discrimination (Marraffini and Sontheimer, 2010). Accumulated evidence points to DNA as the target for CRISPR-mediated degradation in *E. coli* (Brouns et al., 2008) and *S. epidermidis* (Marraffini and Sontheimer, 2008); however, recent studies of *P. furiosus* raise the possibility that some CRISPR-Cas systems target RNA (Hale et al., 2009). This functional difference seen for *P. furiosus* may be due to the presence of a Cas repeat-associated mysterious protein module, which contains six genes, cmr1–cmr6, that are absent in numerous prokaryotes including *E. coli* and *S. epidermidis* (Haft et al., 2005; Makarova et al., 2006). Our own studies suggest that plasmid DNA is the intracellular target of the *E. coli* CRISPR-Cas system; however, whether the Cas enzymes alternatively target the RNA of a factor required for plasmid replication, segregation and/or maintenance cannot be ruled out. Hence, a key issue moving forward is to definitively prove whether Cas enzymes silence DNA and/or RNA in *E. coli*. Based on the observation that mature crRNAs remain associated with *P. furiosus* Cas6 (Carte et al., 2008) and the *E. coli* CasCDE complex (our unpublished observations), we speculate that the same complex may participate directly in target degradation. One possibility for DNA targets is that active replication of phage genome or plasmid DNA allows a window of opportunity for crRNA to base pair with the single-stranded target DNA. Such a mechanism is reminiscent of eukaryotic RNAi, in which Argonaute2 retains one strand of the small interfering RNA (siRNA) as the guide to base pair and cleave target mRNAs (Liu et al., 2004). While the composition of the in vivo crRNP effector complex remains a mystery, our genetic screen showed that Cas1, Cas2, Cas3 and all five Cascade proteins are required for ssTorA–GFP silencing in a Δ*dnaK* strain background. Biochemically only the CasCDE enzymes were required in CRISPR processing but they did not exhibit any measurable DNase activity. Thus, given that *E. coli* Cas1 and Cas2 function upstream of the CRISPR processing and DNA

targeting stages (Brouns et al., 2008), it follows that CasA, CasB and/or Cas3 may participate in the downstream effector stage (Fig. 4.9). We also cannot rule out the participation of other enzymes that may be encoded outside of the cas operon. We are currently investigating whether the CasCDE complex serves as a landmark to attract effector proteins through transient interactions (Fig. 5.9).

5.10 Materials and methods

5.10.1 Bacterial strains, plasmids and growth conditions

Escherichia coli strain BW25113 (*lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$) and single-gene knockout mutants of BW25113 from the Keio collection (Baba et al., 2006) were used for all experiments unless otherwise noted. BW25113 Δ*dnaK* double knockout strains were created by P1*vir* phage transduction. Briefly, kanamycin (Kan)-marked alleles derived from single-gene knockout strains from the Keio collection were transduced in recipient BW25113 Δ*dnaK* cells that had the Kan resistance marker in *dnaK* eliminated as described (Datsenko and Wanner, 2000). The CRISPR deletion strain was made by generating a Kan-marked deletion in BW25113 Δ*dnaK* cells as described (Datsenko and Wanner, 2000). BL21(DE3) [F^{-}*ompT hsdS*$_{B}$ (r_{B^-} m_{B^-}) *gal dcm*λ(DE3)] (Novagen) or T7 Express *lysY* (New England Biolabs) cells were used as hosts for expression of the BaeR and Cas enzymes respectively.

Strains were routinely grown aerobically at 37°C in Luria–Bertani (LB) medium, and antibiotics were supplemented at the following concentrations: Amp, 100 μg ml^{-1}; Kan (5 μg ml^{-1}); spectinomycin (Spec, 50 μg ml^{-1}); tetracycline (Tet, 10 μg ml^{-1}). Protein synthesis was induced when the cells reached an absorbance at 600 nm (Abs_{600}) of ~0.5 by adding 0.2%

arabinose and/or 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG) to the media. For silencing experiments, protein synthesis was induced for 4–6 h prior to characterizing cellular fluorescence or harvesting plasmid DNA, total mRNA and total proteins.

5.10.2 Plasmids

Plasmid pTG encoded the *E. coli* ssTorA signal peptide fused to GFP in pTrc99A (Pérez-Rodríguez et al., 2007) and pTrc–GFP encoded GFP cloned in pTrc99A (Amann and Brosius, 1985). For *dnaK* genetic complementation studies, BW25113 Δ*dnak* cells were transformed with pOFXbad-KJ2, which encodes *dnaK* and *dnaJ* under control of the pBAD promoter (Castanié et al., 1997). For *casE* genetic complementation studies, BW25113 Δ*dnak*Δ*casE* cells were transformed with pOFXbad–casE, which encodes *casE* under control of the pBAD promoter. Synonymous mutations in the ssTorA signal peptide were created using site-directed mutagenesis (QuikChange® II Site-Directed Mutagenesis Kit, Stratagene) according to manufacturer's protocols. Plasmid pssS5–GFP was created by mutating the C57 and C60 nucleotides in *sstorA* to T57 and T60 respectively. Plasmid pss1.5.8–GFP was created by further mutating pssS5–GFP as follows: G99T, G102A, G126C and G129C. Similarly, plasmid pTorA(KK)–GFP was created by mutating the consensus twin-arginine residues in the ssTorA signal peptide to twin-lysine residues using site-directed mutagenesis. For protein purification, CasC, CasE and $CasE^{H20A}$ were PCR amplified from *E. coli* K12 genomic DNA and cloned into pET-21a(+) (Novagen) with an N-terminal His_6-tag. CasA, CasB, CasD, Cas1 and Cas3 were cloned into the pQE-80 vector (Qiagen) with an N-terminal His_6-tag. Cas2 was cloned in pGEX-4T-1 (GE Healthcare) to generate a GST fusion. Additional plasmids used in Cas protein coexpression studies included: CasC cloned in the pCDFDuet-1 vector (Novagen) without any affinity tag, CasD cloned into the

pSUMO vector with an N-terminal His_6-SUMO tag and CasE cloned in pET-29b(+) (Novagen) without any affinity tag. $CasE^{H20A}$ was constructed using Phusion Site-Directed Mutagenesis Kit (New England Biolabs). BaeS and BaeR were cloned in pET-21a(+) (Novagen) with a C-terminal His_6-tag. For *cas* complementation studies, plasmid pWUR399, pWUR402 (Brouns et al., 2008) or pWUR399($CasE^{H20A}$) was cotransformed in strain BL21(DE3) along with pTG and induced using IPTG. Plasmid pWUR399($CasE^{H20A}$) was created by site-directed mutagenesis of plasmid pWUR399.

5.10.3 Generation of transposon library and mapping of insertions

An initial mini-Tn*10* transposon insertion mutant library was generated by tranducing *E. coli* NLL51, an MC4100 *recA*– derivative, to tetracycline resistance using a population of lambda phage that carried mini-Tn*10* inserts as described previously (Takiff et al., 1989). The initial mini-Tn*10* insertion library contained 4.2×10^4 transductants, representing ~10-fold coverage of the genome. This library was then transduced in Δ*dnaK* cells carrying pTG by P1*vir* phage transduction and 3.0×10^4 clones were recovered on LB-Tet/Amp plates, for approximately eightfold coverage of the genome. To isolate highly fluorescent clones, library cells were grown overnight, subcultured and induced for 4–6 h. Induced library cells were diluted 200-fold in 1 ml of phosphate-buffered saline (PBS, pH 7.4) and sorted using a FACSCalibur flow cytometer (Becton Dickinson Biosciences) as described previously (DeLisa and Bentley, 2002). A total of ~1×10^6 cells were examined in 30 min, and ~400 events were collected. Individual clones were re-screened via flow cytometry and a fluorescent plate reader (Synergy HT, BioTek Instruments) for verification of fluorescent phenotypes.

To sequence the insertion mutations in the chromosome of positive clones, a nested PCR strategy was used. This entailed a first PCR reaction using genomic DNA from Δ*dnaK* mini-Tn*10* mutants as template, and a reverse primer (CTTTGGTCACCAACGCTTTTCCCG) specific for the 3′ end of the mini-Tn*10* insert and a forward arbitrary primer (GGCCACGCGTCGACTAGTACNNNNNNNNNNGCTGG) that annealed to the flanking chromosome. A second PCR was then performed using the PCR product from the first reaction as template, along with a forward primer (GGCCACGCGTCGACTAGTAC) that annealed to the constant portion of the arbitrary primer and a reverse primer (CATATGACAAGATGTGTATCCACC) that annealed to an upstream region of the mini-Tn*10* sequence. The resulting amplified DNA was gel purified and sequenced.

5.10.4 Cellular fluorescence analysis

Overnight cultures harbouring GFP-based plasmids were subcultured into fresh LB medium with appropriate antibiotics and protein expression was induced in mid-exponential phase growth. After 4–6 h, cells were harvested by centrifugation and washed once with PBS, pH 7.4. Next, 5 μl of washed cells were diluted into 1 ml of PBS and analysed using a FACSCalibur flow cytometer and CellQuest Pro software. All mean cell fluorescence data from flow cytometric analysis are the average of three replicate experiments (n = 3). Error bars represent the standard error of the mean (s.e.m.). Fluorescence microscopy was performed as described (Kim et al., 2008).

\

5.10.5 RNA and protein expression analysis

Total RNA was isolated from 10 ml of cells using the Qiagen RNeasy Mini Kit. Northern blots were performed as follows: 5 μg of total RNA was separated on a formaldehyde agarose gel, transferred to a positively charged nylon membrane by capillary elution and immobilized by baking the membrane. Digoxigenin-labelled oligonucleotide probes were synthesized using PCR amplification in the presence of digoxigenin-dUTP according to manufacturer's protocol (Roche Applied Sciences). Hybridizations were performed under stringent conditions (55°C in 50% formamide, 0.25 M sodium phosphate pH 7.2, 0.25 M sodium chloride, 7% SDS, 100 μg ml^{-1} fragmented salmon sperm DNA and 5 μg ml^{-1} yeast tRNA). Membranes were washed and exposed to anti-digoxigenin antibody from mouse (diluted 1:10 000) and CDP-Star substrate according to manufacturer's protocol (Roche Applied Sciences). Detection was with X-ray film. Cytoplasmic and periplasmic protein fractions were generated from 10 ml of cells using the cold osmotic shock procedure (Kim et al., 2005). Subcellular fractions were separated by SDS-PAGE using 12% Tris-HCl gels (Bio-Rad), transferred to Immobilon P (Millipore) and blotted using the following primary antibodies: mouse anti-GFP (Sigma; diluted 1:2 000) and mouse anti-GroEL (Sigma; diluted 1:20 000). Secondary antibodies were goat anti-mouse and goat anti-rabbit horseradish peroxidase conjugates (Promega) diluted 1:2500. Signals were detected using enhanced chemiluminescence (ECL) (Amersham). To verify the fractionation quality, membranes were first probed with primary antibodies and, following development, stripped in Tris-buffered saline/2% (w/v) SDS/0.7 M β-mercaptoethanol. Stripped membranes were re-blocked and probed with anti-GroEL antibodies.

5.10.6 Protein purification

For biochemical studies, Cas protein expression vectors were transformed into T7 Express *lysY* host cells (New England Biolabs) and expression was induced at 15°C for 20 h with 0.3 mM IPTG. The harvested cells were lysed by sonication in lysis buffer (50 mM Tris-HCl pH 8.5, 500 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol and 1 mM benzamidine chloride). The supernatant after centrifugation was applied onto 5 ml of Ni-sepharose column (Sigma) for each purification. After protein binding, the column was washed thoroughly with 100 column volumes of lysis buffer followed by 10 volumes of lysis buffer supplemented with 40 mM imidazole. The bound protein was eluted from the column using five column volumes of elution buffer (300 mM imidazole pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol and 1 mM benzamidine chloride). Each purified protein was concentrated and further purified using a Superdex 200 size-exclusion column followed by an ion-exchange column (MonoQ or SourceS) (GE Healthcare). Cas2 was expressed and purified in an N-terminal GST fusion form. All of the purified proteins were analysed by SDS-PAGE and judged to be greater than 95% purity.
To coexpress the CasCE complex, pCDFDuet-1–CasC and pET-21–CasE were cotransformed into T7 Express *lysY* cells. Following coexpression, the CasCE complex was purified on a Ni-sepharose column followed by a Superdex 200 sizing column and a SourceQ ion-exchange column on FPLC as described above. The fact that CasC copurified with His_6-tagged CasE protein despite stringent high-salt wash suggests that these two proteins form a very stable complex. Following a similar strategy, we confirmed that the coexpressed CasC (from pCDFDuet-1–CasC) and CasD (from pQE80–CasD) proteins form a stable complex. In this case, a single His_6-tag was introduced at the N-terminus of the CasD protein.

To test whether a ternary complex formed between CasC, CasD and CasE, we cotransformed pET-29–CasE (encoding CasE without a His_6-tag), pCDFDuet-1–CasC and pQE80–CasD (encoding CasD with an N-terminal His_6-tag) into the host cell T7 Express *lysY*. Following coexpression, the CasCDE complex was purified with a Ni-sepharose column followed by a Superdex 200 column and a SourceQ ion-exchange column on FPLC as described above. An apparent stoichiometry of 2:1:2 was observed for the CasC, CasD and CasE proteins in the ternary complex after the above purification (data not shown). However, after extended incubation at 4°C and heparin column separation, the stoichiometry of the ternary complex shifted to 6:1:1, reflecting the tendency of CasC to oligomerize into most likely a helical hexamer in a dynamic fashion. Although only CasD contains a His_6-tag, CasE and CasC copurifies with CasD under stringent purification conditions, suggesting that these three proteins indeed form a stable ternary complex. Because the CasCDE complex could not be satisfactorily separated from the CasCD complex by size-exclusion or ion-exchange chromatography, another coexpression system was designed in which CasD was cloned into the pSUMO vector with an N-terminal His_6-SUMO tag. The resulting $CasCED_{SUMO}$ complex again copurified on the Ni-sepharose column under stringent purification conditions, and $CasCED_{SUMO}$ and CasCD were separated on Superdex 200 column based on their size difference.

5.10.7 RNase activity assay

The HEX-labelled RNA substrate (5′-GAGUUCCCCGCGCCAGCGGGGAUAAACCG-3′) was purchased from Sigma. RNA and proteins were incubated for 30 min in a buffer containing 20 mM Tris pH 7.5, 20 mM NaCl, 1 mM $MgAc_2$ and 10% glycerol. The reaction mixture was combined with loading buffer at a

ratio of 1:9 and separated on an 18% sequencing gel. The gel was scanned with a Typhoon 9400.

5.10.8 DNase activity assay

The ds-DNase activity was carried out in 20 μl reactions at 37°C for 2 h in a buffer containing 10 mM Tris-HCl pH 7.9, 50 mM NaCl, 10 mM $MgCl_2$ and 1 mM dithiothreitol using 0.05 g l^{-1} supercoiled pTG or pUC19 plasmids as substrates. The reaction was then phenol-extracted, separated on a 0.8% agarose gel and stained with ethidium bromide. The ss-DNase assay was carried out in the same buffer. 5′ Fluorescein-labelled 29 nt DNA oligo (fluorescein-TCGCGCATACCCTGCGCGTCGCCGCC) at a concentration of 100 nM encoding part of spacer 5 was used as substrate to allow fluorescent detection. In a parallel set of experiment, this oligo was heat-annealed with equimolar amount of the complementary strand to assay for the ds-DNase activity under the same conditions. The reaction mixture was separated in 18% sequencing gel, and the fluorescent signals were read using a Typhoon 9400 Imager. All reactions were carried out at 37°C with the addition of 200 nM Cas proteins/complexes.

5.10.9 Plasmid DNA restriction analysis

Plasmid DNA was isolated from an equivalent number cells carrying either pTG and or empty pTrc99A (control) using a Qiagen Miniprep kit. An equal volume of plasmid DNA was digested with the non-unique restriction enzyme RsaI (New England Biolabs) for 4 h at 37°C. After removing the enzyme from the reaction using a PCR cleanup kit (Qiagen), 20 μl of each DNA-containing sample was mixed with loading dye (1:1) and separated on a 1% agarose gel pre-stained with ethidium bromide.

5.10.10 Bacterial MIC and viability analysis

The MIC of bacteria grown was determined by selective plating of cells on Amp. Overnight cultures were diluted 10^5-fold in liquid LB and plated on LB agar supplemented with increasing Amp concentrations and 0.1 mM IPTG. Reported MIC values were the average of three replicate experiments. The number of colony-forming units per millilitre (cfu ml^{-1}) was determined as follows: overnight cultures were subcultured and induced with 1 mM IPTG after reaching an Abs_{600}~0.5. Following 4 h of induction, an equivalent amount of cells (Abs_{600}~1.0) was harvested and plated on LB agar with no antibiotics. The number of cfu ml^{-1} was determined by counting individual colonies on plates after overnight growth. Reported cfu ml^{-1} values were the average of three replicate experiments.

5.10.11 Electrophoretic mobility shift analysis

Phosphorylated or unphosphorylated BaeR was assayed for its ability to bind putative Cas promoter DNA in the vicinity of the *casA* and *cas3* genes by electrophoretic mobility shift analysis. Putative Cas promoter DNA in the vicinity of the *casA* and *cas3* genes cloned from BW25113 wt genomic DNA by PCR using the following primers: *casA* forward: 5′TAAACCGCTTTTAAAACCACCACCAAAACCT-3′, *casA* reverse: 5′GATGGATGGGTCTGGCAGGG-3′; *cas3* forward: GGCATATATATTTAAAAGGTTCCATTAATAGCCTCCCTGTTTTTTTAGTA, *cas3* reverse: GGTTACACGAAGGGTAAATATTGCCGGACAAATT. Control DNA fragments corresponding to the *acrD* (−1 to −276 bp) promoter region were similarly prepared by PCR with BW25113 genomic DNA as a template. Separately, BaeR-His_6 was expressed in BL21(DE3) cells from pET-21–BaeR and purified using a Ni-sepharose column (Qiagen). Purified BaeR-His_6 was

phosphorylated using 50 mM acetylphosphate in a buffer containing 100 mM Tris (pH 7.4), 10 mM $MgCl_2$ and 125 mM KCl. The protein was incubated in this buffer for 1 h at 30°C. Next, 0.15 pmol of DNA was mixed with varying amounts (9.2, 11.5, 16.1, 24.2 and 32.25 μM) of the phosphorylated protein in a 20 μl reaction containing 100 mM Tris (pH 7.4), 10 mM $MgCl_2$, 100 mM KCl, 10% glycerol and 2 mM dithiothreitol. The reaction mixture was incubated for 20 min at room temperature and the samples were electrophoresed on a 10% Tris-Borate-EDTA (TBE) gel (Bio-Rad) in TBE buffer at 4°C. The gel was then soaked in a 1:10 000 dilution of SYBR-Gold stain (Invitrogen) and the DNA was imaged under UV light using a Gel Doc 2000 (Bio-Rad).

5.10.12 Analysis of gene expression by qRT-PCR

Quantitative reverse transcription polymerase chain reaction analysis of *cas*, *spy*, and *sstora-gfp* gene transcript abundance was performed similar to that described elsewhere (Westra et al., 2010). Briefly, cDNA was synthesized using Omniscript RT Kit (Qiagen) from RNA extracted using the RNeasy Mini Kit (Qiagen) and DNase I-treated using RNase-free DNase Set (Qiagen). The qRT-PCR reactions were performed using TaqMan Universal PCR Master Mix (Applied Biosystems) according to manufacturer's instructions. The PCR reactions were performed on an ABS Prism 7000 Sequence Detection System (Applied Biosystems) and analysed using ABS Prism 7000 SDS software (Applied Biosystems). Fold change of gene transcription was calculated using the relative quantification method with *rrsB* endogenous control. The gene transcript abundance of *cas*, *spy* and *sstorA–gfp* in *E. coli* BW25113 wt cells harbouring pTG were used as calibrators. All PCR reactions were performed in triplicate.

CHAPTER 6

PROSPECTIVE APPLICATIONS IN ENGINEERING EXTRACELLULAR SECRETION

6.0 Introduction

The work presented in this dissertation challenges the status quo that non-pathogenic *E. coli* is a poor secretor of proteins. There is an increasing body of evidence presented here and in other studies that suggest that *E. coli* encodes secretory mechanisms capable of high-titer production of heterologous proteins in the extracellular medium. Perhaps the most outstanding of these secretory mechanisms is the YebF secretion pathway, which promoted much greater secretion than any of the other single-domain *E. coli* secretion systems such as OsmY, OmpF, and OmpA in our studies. Furthermore, the YebF pathway's tolerance to secrete heterologous proteins with diverse functional and structural properties makes it a powerful tool for many biotechnology applications that benefit from or require extracellular biosynthesis of proteins, in particular consolidated bioprocessing which requires extracellular enzymes to deconstruct plant biomass.

Due in part to *E. coli*'s genetic tractability, we were able to reprogram cells for increased secretion levels of YebF by employing a FlAsH-based high-throughput screening platform. These engineered strains demonstrated increased secretion levels of both YebF and YebF-fusion proteins. In addition to the FlAsH-based screen, we engineered a genetic selection whereby the non-secretory phenotype was lethal, which further streamlines the identification of rare genetic events that effect secretion activity. Taken together, our new approach for engineering extracellular secretion pathways opens the door to creating superior expression hosts for the

extracellular biosynthesis of heterologous proteins. In this chapter, I will describe some future work that builds upon the tools developed in the previous chapters of this thesis.

6.1 Cellulosomes

Plant pathogens have evolved two common mechanisms to deconstruct plant biomass during host invasion. One mechanism is to secrete PCWDEs freely to the extracellular milieu, typically by T2SS or other secretory mechanism. Once secreted, these enzymes are no longer associated with the host cell, and this free-enzyme secretion strategy is typically employed by aerobic organisms. An alternative strategy, more commonly found in anaerobic bacteria such as *Clostridium thermocellum*, is to tether enzymes to the cell surface as large (megaDalton) extracellular complexes called cellulosomes. The assembly of cellulosomes on the cell surface offers distinct advantages over free-enzyme systems. For example, the cellulosome is sandwiched between the host cell and the plant biomass material, allowing easier access of released nutrients to the host. Another advantage is that cellulosomes give the host the ability to control the spatial organization of enzymes and their stoichiometry. The result is an enzyme channeling effect that can produce synergies among enzymes not normally achieved in free enzyme systems (Fontes and Gilbert, 2010; Delisa and Conrado, 2009).

The construction of cellulosomes centers on a non-cellulolytic scaffoldin that is anchored in the host outer membrane. The scaffoldin is decorated with cohesion domains, which have affinity for corresponding dockerin domains that are encoded in the cellulases, hemicellulases, and carbohydrate-bind modules (CBMs). The CBMs play a critical role in mediating protein-carbohydrate interactions and targeting the catalytic domains of the cellulosome to the plant substrate (Fontes and Gilbert, 2010). The result is a highly modular cellulolytic complex

whereby catalytic enzymes are mounted on a non-catalytic scaffoldin via their dockerin-cohesion domains. The modular nature of cellulosomes can be exploited in the laboratory to create small-scale 'designer' cellulosomes. For example, engineered miniature cellulosomes have successfully been expressed in *S. cervisiae* (Hyeon et al., 2010; Tsai and Goyal, 2010) and also *Bacillus subtilis* (Anderson et al., 2011). Engineering mini-cellulosomes in these strains represents a step forward in engineering a consolidated bioprocessing microbe capable of high level ethanol production directly from plant biomass.

A requirement for the assembly of mini-cellulosomes is the extracellular production of cellulases that become mounted on surface-displayed scaffoldins. In a previous study, bacterial mini-cellulosomes were constructed by adding purified cellulases to scaffoldin-expressing cells, or by co-expressing the enzymes with the scaffoldin (Anderson et al., 2011). As a result, a large portion of protein (~70%) was not anchored to the cell wall (Anderson et al., 2011). An alternate strategy for assembling mini-cellulosomes is to target the cellulases for extracellular secretion using a dedicated secretion system to allow them to self-assemble on scaffoldins displayed on the cell exterior. A particularly attractive secretion system for this purpose is the YebF system. Our studies have shown that YebF is capable of carrying at least four different cellulase enzymes by a single-host (Fig. 3.1). It is also possible that high secretion levels are necessary to saturate the dockerin site(s) decorated on the scaffoldins. To address this issue, the secretion of YebF-cellulase could be combined with a host that has been reprogrammed for hypersecretion of YebF proteins to achieve secretion titers suitable for efficient scaffoldin-mediated surface display of these enzymes. The *E. coli* strain BW25113 Δ*nlpD* is an attractive host for this purpose because it yielded a high level of YebF secretion by FlASH analysis (Fig. 3.3), and the integrity of its outer membrane did not appear to be compromised (Table 3.4).

6.2 Optimizing the secretion of specific POI

In Chapter 4, we describe the establishment of an engineered genetic selection for studying extracellular secretion pathways. The selection functions by the activity of the BLIP protein which inhibits the β-lactamase enzyme in the periplasmic compartment. When BLIP is fused to a secretory protein, such as YebF, it is exported out of the periplasm and into the culture medium, and as a result cells are more resistant to β-lactam antibiotics. This gives rise to varying resistance phenotypes that can be exploited to screen large or multiple genetic libraries to identify rare genetic events that affect secretion efficiency. For example, a plasmid encoding the YebF-BLIP selection system can be transformed into a library of *E. coli* strains carrying non-essential gene deletions (known as the Keio library) (Baba et al., 2006) to identify *trans*-acting genetic factors that influence export of the YebF-BLIP fusion protein. Furthermore, this resistance phenotype makes it possible to screen large genetic libraries that are typically needed in directed evolution experiments. Thus, it should be possible to reprogram the YebF secretion machinery by screening error-prone PCR libraries of the OmpF, OmpC, and OmpX proteins to identify variants with improved activity.

These strategies are not limited only to the YebF-BLIP fusion. It should be possible to employ the same strategies for engineering the secretion of virtually any POI, such as cellulases or other heterologous proteins. This can be done my inserting the POI between the YebF-BLIP fusion to create a YebF-POI-BLIP construct. A YebF-POI-BLIP construct should in principle be amenable to screening the same libraries as described above with the YebF-BLIP construct. Thus, it would be interesting to compare the genetic mutations that are isolated with the YebF-BLIP

versus the YebF-POI-BLIP construct. This leads to the tantalizing possibility that there are common mutations that affect the secretion of all YebF-fusion proteins.

6.3 Optimizing co-secretion from a single host

Co-secretion of different proteins is required in some biomolecular engineering contexts. This is particularly true for the construction of cellulosomes in consolidated bioprocessing platforms because efficient deconstruction of plant biomass requires multiple cellulase enzymes with different activities (Liao et al., 2011). Our studies have made it possible to engineer a single host for co-secretion of more than one protein into the extracellular environment. Furthermore, we have uncovered a method for reprogramming these cells to enhance the co-secretion efficiency of multiple proteins together. There are two critical observations in Chapter 3 that can be exploited to enhance the simultaneous co-secretion of heterologous proteins. First, while the YebF pathway was known to efficiently carry heterologous proteins to the culture supernatant (Zhang et al., 2006), it was unknown whether the YebF pathway could accommodate the secretion of more than one protein from a single host. Our studies indicate that the YebF pathway can secrete at least four different proteins. This was evident by immunoblot analysis that showed the presence of YebF-Cel3A, YebF-Cel5B, YebF-Cel6A, and YebF-Cel9A in the culture supernatant when expressed from a single plasmid (Fig. 3.1b). The second observation is that cells secreting these four proteins can grow on M9 minimal medium containing CMC as a sole carbon source, and this growth phenotype minimally required YebF-Cel3A and YebF-Cel5B (Table 3.1.2).

These two observations combined can be exploited to reprogram cells for enhanced co-secretion of more than one protein by a single host cell. For example, the viability phenotype on

CMC creates a powerful method for screening combinatorial libraries to identify rare genetic events that enhance the secretion of both YebF-Cel3A and YebF-Cel5B together. By selecting cell growth on CMC, genome-wide transposon insertion libraries can be screened in this system to find *trans*-acting factors that enhance the secretion of YebF-Cel3A and YebF-Cel5B together. Alternatively, error-prone PCR libraries can be created in the YebF domains of YebF-Cel3A and YebF-Cel5B to find variants that lead to higher secretion titers. It would be interesting to compare the genetic mutations that arise in engineering co-secretion of YebF-Cel3A and YebF-Cel5B with the mutations that arise from screening transposon-insertion libraries using our YebF-FT system because these studies may lead to common mutations that enhance secretion of all proteins. Finally, it is possible that overexpression of YebF proteins can saturate the YebF secretion system. Thus, the YebF carrier domain can be swapped with other secretion domains, such as OsmY, to test whether these systems alone or in combination with YebF can enhance secretion efficiency.

REFERENCES

Alaimo, C., Catrein, I., Morf, L., Marolda, C.L., Callewaert, N., Valvano, M.A., Feldman, M.F., and Aebi, M. (2006). Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. Embo J *25*, 967–976.

Altaras, N.E., and Cameron, D.C. (1999). Metabolic engineering of a 1,2-propanediol pathway in *Escherichia coli*. Appl Environ Microbiol *65*, 1180–1185.

Altaras, N.E., and Cameron, D.C. (2000). Enhanced production of (R)-1,2-propanediol by metabolically engineered *Escherichia coli*. Biotechnol Prog *16*, 940–946.

Amann, E., and Brosius, J. (1985). "ATG vectors' for regulated high-level expression of cloned genes in *Escherichia coli*. Gene *40*, 183–190.

Amann, E., Ochs, B., and Abel, K.J. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene *69*, 301–315.

Anderson, T.D., Robson, S.A., Jiang, X.W., Malmirchegini, G.R., Fierobe, H.-P., Lazazzera, B.A., and Clubb, R.T. (2011). Assembly of minicellulosomes on the surface of *Bacillus subtilis*. Appl Environ Microbiol *77*, 4849–4858.

Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim. Biophys. Acta *1473*, 4–8.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol *2*, 2006.0008–.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. Nat Biotechnol *22*, 1399–1408.

Baranova, N., and Nikaido, H. (2002). The baeSR two-component regulatory system activates transcription of the yegMNOB (mdtABCD) transporter gene cluster in *Escherichia col*i and increases its resistance to novobiocin and deoxycholate. J Bacteriol *184*, 4168–4176.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Ben-Dor, S., Esterman, N., Rubin, E., and Sharon, N. (2004). Biases and complex patterns in the residues flanking protein *N*-glycosylation sites. Glycobiology *14*, 95–101.

Blocker, A.A., Komoriya, K.K., and Aizawa, S.-I.S. (2003). Type III secretion systems and bacterial flagella: insights into their function from structural similarities. Proc Natl Acad Sci USA *100*, 3027–3030.

Bokinsky, G., Peralta-Yahya, P.P., George, A., Holmes, B.M., Steen, E.J., Dietrich, J., Soon Lee, T., Tullman-Ercek, D., Voigt, C.A., Simmons, B.A., et al. (2011). Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. Proc Natl Acad Sci USA *108*, 19949–19954.

Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S.D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G.D., et al. (2004). Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat Biotechnol *22*, 1554–1558.

Bos, M.P., Robert, V., and Tommassen, J. (2007). Biogenesis of the Gram-Negative Bacterial Outer Membrane. Annu Rev Microbiol *61*, 191–214.

Bouley, J., Condemine, G., and Shevchik, V.E. (2001). The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II out pathway of *Erwinia chrysanthemi*. J Mol Biol *308*, 205–219.

Boyd, A.P., Grosdent, N., Tötemeyer, S., Geuijen, C., Bleves, S., Iriarte, M., Lambermont, I., Octave, J.N., and Cornelis, G.R. (2000). *Yersinia enterocolitica* can deliver Yop proteins into a wide range of cell types: development of a delivery system for heterologous proteins. Eur. J. Cell Biol. *79*, 659–671.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev *22*, 3489–3496.

Castanié, M.P., Bergès, H., Oreglia, J., Prère, M.F., and Fayet, O. (1997). A set of pBR322-compatible plasmids allowing the testing of chaperone-assisted folding of proteins overexpressed in *Escherichia coli*. Anal Biochem *254*, 150–152.

Chen, W., and Helenius, A. (2000). Role of ribosome and translocon complex during folding of influenza hemagglutinin in the endoplasmic reticulum of living cells. Mol. Biol. Cell *11*, 765–772.

Collins, C.H.C., Leadbetter, J.R.J., and Arnold, F.H.F. (2006). Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. Nat Biotechnol *24*, 708–712.

Cornelis, G.R. (2006). The type III secretion injectisome. Nat Rev Microbiol *4*, 811–825.

Darwin, A.J. (2005). The phage-shock-protein response. Mol Microbiol *57*, 621–628.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA *97*, 6640–6645.

DeBoy, R.T., Mongodin, E.F., Fouts, D.E., Tailford, L.E., Khouri, H., Emerson, J.B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H.J., et al. (2008). Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus. J Bacteriol *190*, 5455–5463.

DeLisa, M.P., and Bentley, W.E. (2002). Bacterial autoinduction: looking outside the cell for new metabolic engineering targets. Microb Cell Fact *1*, 5.

Delisa, M.P., and Conrado, R.J. (2009). Synthetic metabolic pipelines. Nat Biotechnol 27, 728–729.

DeLisa, M.P., Lee, P., Palmer, T., and Georgiou, G. (2004). Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway. J Bacteriol *186*, 366–373.

DeLisa, M.P., Samuelson, P., Palmer, T., and Georgiou, G. (2002). Genetic analysis of the twin arginine translocator secretion pathway in bacteria. J Biol Chem *277*, 29825–29831.

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

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol *190*, 1390–1400.

Díez-Villaseñor, C.C., Almendros, C.C., García-Martínez, J.J., and Mojica, F.J.M.F. (2010). Diversity of CRISPR loci in *Escherichia coli*. CORD Conference Proceedings *156*, 1351–1361.

Donath, M.J., Dominguez, M.A., and Withers, S.T. (2011). Development of an automated platform for high-throughput P1-phage transduction of *Escherichia coli*. J Lab Autom *16*, 141–147.

Ellis, R.J., and Hartl, F.U. (1996). Protein folding in the cell: competing models of chaperonin function. Faseb J *10*, 20–26.

Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., and Aebi, M. (2005). Engineering *N*-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. Proc Natl Acad Sci USA *102*, 3016–3021.

Fernandez, S., Palmer, D.R., Simmons, M., Sun, P., Bisbing, J., McClain, S., Mani, S., Burgess, T., Gunther, V., and Sun, W. (2007). Potential role for Toll-like receptor 4 in mediating *Escherichia coli* maltose-binding protein activation of dendritic cells. Infect Immun *75*, 1359–1363.

Fisher, A.C., Haitjema, C.H., Guarino, C., Celik, E., Endicott, C.E., Reading, C.A., Merritt, J.H., Ptak, A.C., Zhang, S., and DeLisa, M.P. (2011). Production of secretory and extracellular *N*-linked glycoproteins in *Escherichia coli*. Appl Environ Microbiol *77*, 871–881.

Fisher, A.C., Kim, W., and DeLisa, M.P. (2006). Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. Protein Sci *15*, 449–458.

Fontes, C.M.G.A., and Gilbert, H.J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu. Rev. Biochem. *79*, 655–681.

Francetic, O., and Pugsley, A.P. (1996). The cryptic general secretory pathway (*gsp*) operon of *Escherichia coli* K-12 encodes functional proteins. J Bacteriol *178*, 3544–3549.

Francetic, O., Badaut, C., Rimsky, S., and Pugsley, A.P. (2000a). The ChiA (YheB) protein of *Escherichia coli* K-12 is an endochitinase whose gene is negatively controlled by the nucleoid-structuring protein H-NS. Mol Microbiol *35*, 1506–1517.

Francetic, O., Belin, D., Badaut, C., and Pugsley, A.P. (2000b). Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion. Embo J *19*, 6697–6703.

Francisco, J.A., and Georgiou, G. (1994). The expression of recombinant proteins on the external surface of *Escherichia coli*. Biotechnological applications. Ann. N. Y. Acad. Sci. *745*, 372–382.

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

Gardner, J.G., and Keating, D.H. (2010). Requirement of the type II secretion system for utilization of cellulosic substrates by *Cellvibrio japonicus*. Appl Environ Microbiol *76*, 5079–5087.

Georgiou, G. (1988). Optimizing the Production of Recombinant Proteins in Microorganisms. Aiche J *34*, 1233–1248.

Gottesman, S.S. (1996). Proteases and their targets in *Escherichia coli*. Genetics *30*, 465–506.

Griffin, B.A., Adams, S.R., and Tsien, R.Y. (1998). Specific covalent labeling of recombinant protein molecules inside live cells. Science *281*, 269–272.

Grissa, I., Vergnaud, G., and Pourcel, C. (2008). CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. Nucleic Acids Res *36*, W145–W148.

Grissa, I.I., Vergnaud, G.G., and Pourcel, C.C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. CORD Conference Proceedings *35*, W52–W57.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comp Biol *1*, e60–e60.

Hale, C., Kleppe, K., Terns, R.M., and Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. RNA *14*, 2572–2579.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945–956.

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

Haurwitz, R.E.R., Jinek, M.M., Wiedenheft, B.B., Zhou, K.K., and Doudna, J.A.J. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. CORD Conference Proceedings *329*, 1355–1358.

He, S.Y., Lindeberg, M., Chatterjee, A.K., and Collmer, A. (1991). Cloned *Erwinia chrysanthemi* out genes enable *Escherichia coli* to selectively secrete a diverse family of heterologous proteins to its milieu. Proc Natl Acad Sci USA *88*, 1079–1083.

Helenius, A., and Aebi, M. (2001). Intracellular functions of *N*-linked glycans. Science *291*, 2364–2369.

Helenius, A., and Aebi, M. (2004). Roles of *N*-linked glycans in the endoplasmic reticulum. Annu. Rev. Biochem. *73*, 1019–1049.

Hirakawa, H., Inazumi, Y., Masaki, T., Hirata, T., and Yamaguchi, A. (2005). Indole induces the expression of multidrug exporter genes in *Escherichia coli*. Mol Microbiol *55*, 1113–1126.

Hodgkinson, J.L., Horsley, A., Stabat, D., Simon, M., Johnson, S., da Fonseca, P.C.A., Morris, E.P., Wall, J.S., Lea, S.M., and Blocker, A.J. (2009). Three-dimensional reconstruction of the *Shigella* T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout (vol 16, pg 477, 2009). Nature Structural & Molecular Biology *16*

Hou, B., Frielingsdorf, S., and Klösgen, R.B. (2006). Unassisted membrane insertion as the initial step in DeltapH/Tat-dependent protein transport. J Mol Biol *355*, 957–967.

Hyeon, J.-E., Yu, K.-O., Suh, D.J., Suh, Y.-W., Lee, S.E., Lee, J., and Han, S.O. (2010). Production of minicellulosomes from *Clostridium cellulovorans* for the fermentation of cellulosic ethanol using engineered recombinant *Saccharomyces cerevisiae*. Fems Microbiol Lett *310*, 39–47.

Ignatova, Z., and Gierasch, L.M. (2004). Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. Proc Natl Acad Sci USA *101*, 523–528.

Jarboe, L.R., Grabar, T.B., Yomano, L.P., Shanmugan, K.T., and Ingram, L.O. (2007). Development of ethanologenic bacteria. Adv Biochem Eng Biot *108*, 237–261.

Jeoh, T., Michener, W., Himmel, M.E., Decker, S.R., and Adney, W.S. (2008). Implications of cellobiohydrolase glycosylation for use in biomass conversion. Biotechnol Biofuels *1*, 10.

Joung, J.K., Ramm, E.I., and Pabo, C.O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. Proc Natl Acad Sci USA *97*, 7382–7387.

Jyot, J.J., Balloy, V.V., Jouvion, G.G., Verma, A.A., Touqui, L.L., Huerre, M.M., Chignard, M.M., and Ramphal, R.R. (2011). Type II secretion system of *Pseudomonas aeruginosa*: in vivo evidence of a significant role in death due to lung infection. J Infect Dis *203*, 1369–1377.

Kazemi-Pour, N., Condemine, G., and Hugouvieux-Cotte-Pattat, N. (2004). The secretome of the plant pathogenic bacterium *Erwinia chrysanthemi*. Proteomics *4*, 3177–3186.

Kelly, J., Jarrell, H., Millar, L., Tessier, L., Fiori, L.M., Lau, P.C., Allan, B., and Szymanski, C.M. (2006). Biosynthesis of the *N*-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. J Bacteriol *188*, 2427–2434.

Kim, J.-Y., Doody, A.M., Chen, D.J., Cremona, G.H., Shuler, M.L., Putnam, D., and DeLisa, M.P. (2008). Engineered Bacterial Outer Membrane Vesicles with Enhanced Functionality. J Mol Biol *380*, 51–66.

Kim, J.-Y., Fogarty, E.A., Lu, F.J., Zhu, H., Wheelock, G.D., Henderson, L.A., and DeLisa, M.P. (2005). Twin-arginine translocation of active human tissue plasminogen activator in *Escherichia coli*. Appl Environ Microbiol *71*, 8451–8459.

Kleanthous, C. (2010). Swimming against the tide: progress and challenges in our understanding of colicin translocation. Nat Rev Microbiol *8*, 843–848.

Konjufca, V., Wanda, S.-Y., Jenkins, M.C., and Curtiss, R. (2006). A recombinant attenuated *Salmonella enterica* serovar Typhimurium vaccine encoding *Eimeria acervulina* antigen offers protection against *E. acervulina* challenge. Infect Immun *74*, 6785–6796.

Korotkov, K.V., Sandkvist, M., and Hol, W.G.J. (2012). The type II secretion system: biogenesis, molecular architecture and mechanism. Nat Rev Microbiol *10*, 336–351.

Kotzsch, A., Vernet, E., Hammarström, M., Berthelsen, J., Weigelt, J., Gräslund, S., and Sundström, M. (2011). A secretory system for bacterial production of high-profile protein targets. Protein Sci *20*, 597–609.

Kowarik, M., Numao, S., Feldman, M.F., Schulz, B.L., Callewaert, N., Kiermaier, E., Catrein, I., and Aebi, M. (2006a). *N*-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. Science *314*, 1148–1150.

Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., et al. (2006b). Definition of the bacterial *N*-glycosylation site consensus sequence. Embo J *25*, 1957–1966.

Kumar, A., Hajjar, E., Ruggerone, P., and Ceccarelli, M. (2010). Structural and dynamical properties of the porins OmpF and OmpC: insights from molecular simulations. J. Phys.: Condens. Matter *22*, 454125.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol *8*, R61–R61.

Laurent, N., Voglmeir, J., and Flitsch, S.L. (2008). Glycoarrays--tools for determining protein-carbohydrate interactions and glycoenzyme specificity. Chem Commun (Camb) 4400–4412.

Liao, H., Zhang, X.-Z., Rollin, J.A., and Zhang, Y.-H.P. (2011). A minimal set of bacterial cellulases for consolidated bioprocessing of lignocellulose. Biotechnology Journal *6*, 1409–1418.

Lillestøl, R.K., Shah, S.A., Brügger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Mol Microbiol *72*, 259–272.

Lindeberg, M. (1992). Analysis of eight out genes in a cluster required for pectic enzyme secretion by *Erwinia chrysanthemi*: sequence comparison with secretion genes from other gram-…. J Bacteriol.

Lindeberg, M., Salmond, G., and Collmer, A. (1996). Complementation of deletion mutations in a cloned functional cluster of *Erwinia chrysanthemi* out genes with *Erwinia carotovora* out homologues reveals OutC and OutD as candidate gatekeepers of species-specific secretion of proteins via the type II pathway. Mol Microbiol *20*, 175–190.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.-J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437–1441.

Lodinová-Žádníková, R., Sonnenborn, U., and Tlaskalová, H. (1998). Probiotics and *E. coli* infections in man. Vet Q *20*, 78–81.

Lynd, L.R., Laser, M.S., Bransby, D., Dale, B.E., Davison, B., Hamilton, R., Himmel, M., Keller, M., McMillan, J.D., Sheehan, J., et al. (2008). How biotech can transform biofuels. Nat Biotechnol *26*, 169–172.

MacRitchie, D.M., Buelow, D.R., Price, N.L., and Raivio, T.L. (2008). Two-component signaling and gram negative envelope stress response systems. Adv. Exp. Med. Biol. *631*, 80–110.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct *1*, 7–7.

Malo, N., Hanley, J.A., Cerquozzi, S., Pelletier, J., and Nadon, R. (2006). Statistical practice in high-throughput screening data analysis. Nat Biotechnol *24*, 167–175.

Mansell, T.J., Linderman, S.W., Fisher, A.C., and DeLisa, M.P. (2010). A rapid protein folding assay for the bacterial periplasm. Protein Sci *19*, 1079–1090.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843–1845.

Marraffini, L.A.L., and Sontheimer, E.J.E. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. CORD Conference Proceedings *463*, 568–571.

Martin, B.R., Giepmans, B.N.G., Adams, S.R., and Tsien, R.Y. (2005). Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. Nat Biotechnol *23*, 1308–1314.

McBroom, A.J., Johnson, A.P., Vemulapalli, S., and Kuehn, M.J. (2006). Outer membrane vesicle production by *Escherichia coli* is independent of membrane instability. J Bacteriol *188*, 5385–5392.

Michaelis, S.S., Chapon, C.C., D'Enfert, C.C., Pugsley, A.P.A., and Schwartz, M.M. (1985). Characterization and expression of the structural gene for pullulanase, a maltose-inducible secreted protein of *Klebsiella pneumonia*e. J Bacteriol *164*, 633–638.

Model, P., Jovanovic, G., and Dworkin, J. (1997). The *Escherichia coli* phage-shock-protein (psp) operon. Mol Microbiol *24*, 255–261.

Nandakumar, M.P., Cheung, A., and Marten, M.R. (2006). Proteomic analysis of extracellular proteins from *Escherichia coli* W3110. *5*, 1155–1161.

Ni, Y., and Chen, R. (2009). Extracellular recombinant protein production from *Escherichia coli*. Biotechnol Lett *31*, 1661–1670.

Oresnik, I.J.I., Ladner, C.L.C., and Turner, R.J.R. (2001). Identification of a twin-arginine leader-binding protein. Mol Microbiol *40*, 323–331.

Paradkar, A.S., Aidoo, K.A., Wong, A., and Jensen, S.E. (1996). Molecular analysis of a beta-lactam resistance gene encoded within the cephamycin gene cluster of *Streptomyces clavuligerus*. J Bacteriol *178*, 6266–6274.

Perez-Rodriguez, R., Haitjema, C., Huang, Q., Nam, K.H., Bernardis, S., Ke, A., and DeLisa, M.P. (2011). Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. Mol Microbiol *79*, 584–599.

Pérez-Melgosa, M., Ochs, H.D., Linsley, P.S., Laman, J.D., van Meurs, M., Flavell, R.A., Ernst, R.K., Miller, S.I., and Wilson, C.B. (2001). Carrier-mediated enhancement of cognate T cell help: the basis for enhanced immunogenicity of meningococcal outer membrane protein polysaccharide conjugate vaccine. Eur J Immunol *31*, 2373–2381.

Pérez-Rodríguez, R., Fisher, A.C., Perlmutter, J.D., Hicks, M.G., Chanal, A., Santini, C.-L., Wu, L.-F., Palmer, T., and DeLisa, M.P. (2007). An essential role for the DnaK molecular chaperone in stabilizing over-expressed substrate proteins of the bacterial twin-arginine translocation pathway. J Mol Biol *367*, 715–730.

Prehna, G., Zhang, G., Gong, X., Duszyk, M., Okon, M., McIntosh, L.P., Weiner, J.H., and Strynadka, N.C.J. (2012). A Protein Export Pathway Involving *Escherichia coli* Porins. Structure.

Pugsley, A.P. (1993). The complete general secretory pathway in gram-negative bacteria. Microbiol Rev *57*, 50–108.

Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol *75*, 1495–1512.

Raffa, R.G., and Raivio, T.L. (2002). A third envelope stress signal transduction pathway in *Escherichia coli*. Mol Microbiol *45*, 1599–1611.

Rao, S.S., Hu, S.S., McHugh, L.L., Lueders, K.K., Henry, K.K., Zhao, Q.Q., Fekete, R.A.R., Kar, S.S., Adhya, S.S., and Hamer, D.H.D. (2005). Toward a live microbial microbicide for HIV: commensal bacteria secreting an HIV fusion inhibitor peptide. Proc Natl Acad Sci USA *102*, 11993–11998.

Rüssmann, H., Shams, H., Poblete, F., Fu, Y., n, J.E.G., and Donis, R.O. (1998). Delivery of epitopes by the *Salmonella* type III secretion system for vaccine development. Science *281*, 565–568.

Rêgo, A.T., Chandran, V., and Waksman, G. (2010). Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. Biochem J *425*, 475–488.

Richter, S., and Brüser, T. (2005). Targeting of unfolded PhoA to the TAT translocon of *Escherichia coli*. J Biol Chem *280*, 42723–42730.

Rinas, U., and Hoffmann, F. (2004). Selective leakage of host-cell proteins during high-cell-density cultivation of recombinant and non-recombinant *Escherichia coli*. Biotechnol Prog *20*, 679–687.

Rogers, P.L., Jeon, Y.J., Lee, K.J., and Lawford, H.G. (2007). *Zymomonas mobilis* for fuel ethanol and higher value products. Adv Biochem Eng Biotechnol *108*, 263–288.

Sanders, C., Wethkamp, N., and Lill, H. (2001). Transport of cytochrome c derivatives by the bacterial Tat protein translocation system. Mol Microbiol *41*, 241–246.

Sandkvist, M., Michel, L.O., Hough, L.P., Morales, V.M., Bagdasarian, M., Koomey, M., DiRita, V.J., and Bagdasarian, M. (1997). General secretion pathway (*eps*) genes required for toxin secretion and outer membrane biogenesis in *Vibrio cholerae*. J Bacteriol *179*, 6994–7003.

Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315–1317.

Sonti, R. (2005). Bacterial type two secretion system secreted proteins: Double-edged swords for plant pathogens. Mol Plant Microbe in *18*, 891–898.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol *6*, 181–186.

Stemmer, W.P. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*, 389–391.

Szymanski, C.M., and Wren, B.W. (2005). Protein glycosylation in bacterial mucosal pathogens. Nat Rev Microbiol *3*, 225–237.

Szymanski, C.M., Yao, R., Ewing, C.P., Trust, T.J., and Guerry, P. (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol Microbiol *32*, 1022–1030.

Takiff, H.E.H., Chen, S.M.S., and Court, D.L.D. (1989). Genetic analysis of the rnc operon of *Escherichia coli*. J Bacteriol *171*, 2581–2590.

Tang, T.-H., Bachellerie, J.-P., Rozhdestvensky, T., Bortolin, M.-L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Hüttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. Proc Natl Acad Sci USA *99*, 7536–7541.

Tauschek, M.M., Gorrell, R.J.R., Strugnell, R.A.R., and Robins-Browne, R.M.R. (2002). Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of *Escherichia coli*. Proc Natl Acad Sci USA *99*, 7066–7071.

Tsai, S.-L., and Goyal, G. (2010). Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. Appl Environ Microbiol *76*, 7514–7520.

Usón, I., Patzer, S.I., Rodríguez, D.D., Braun, V., and Zeth, K. (2012). The crystal structure of the dimeric colicin M immunity protein displays a 3D domain swap. J Struct Biol *178*, 45–53.

Verez-Bencomo, V., Fernández-Santana, V., Hardy, E., Toledo, M.E., Rodríguez, M.C., Heynngnezz, L., Rodriguez, A., Baly, A., Herrera, L., Izquierdo, M., et al. (2004). A synthetic conjugate polysaccharide vaccine against *Haemophilus influenzae* type b. Science *305*, 522–525.

Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., et al. (2002). *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science *298*, 1790–1793.

Waraho, D., and DeLisa, M.P. (2009). Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. Proc Natl Acad Sci USA *106*, 3692–3697.

Waters, L.S., and Storz, G. (2009). Regulatory RNAs in Bacteria. Cell *136*, 615–628.

Westendorf, A.M.A., Gunzer, F.F., Deppenmeier, S.S., Tapadar, D.D., Hunger, J.K.J., Schmidt, M.A.M., Buer, J.J., and Bruder, D.D. (2005). Intestinal immunity of *Escherichia coli* NISSLE 1917: a safe carrier for therapeutic molecules. FEMS Immunol Med Microbiol *43*, 373–384.

Westra, E.R., Pul, U., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., et al. (2010). H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. Mol Microbiol *77*, 1380–1393.

Whitley, P.P., Nilsson, I.M.I., and Heijne, von, G.G. (1996). A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain. J Biol Chem *271*, 6241–6244.

Widmaier, D.M., Tullman-Ercek, D., Mirsky, E.A., Hill, R., Govindarajan, S., Minshull, J., and Voigt, C.A. (2009). Engineering the *Salmonella* type III secretion system to export spider silk monomers. Mol Syst Biol *5*, 309–.

Wilson, D.B. (2009). Cellulases and biofuels. Curr Opin Biotechnol *20*, 295–299.

Xia, X.-X., Han, M.-J., Lee, S.Y., and Yoo, J.-S. (2008). Comparison of the extracellular proteomes of *Escherichia coli* B and K-12 strains during high cell density cultivation. Proteomics *8*, 2089–2103.

Yamamoto, K., Ogasawara, H., and Ishihama, A. (2008). Involvement of multiple transcription factors for metal-induced spy gene expression in *Escherichia coli*. J Biotechnol *133*, 196–200.

You, L., and Arnold, F.H. (1996). Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. Protein Eng *9*, 77–83.

Young, N.M.N., Brisson, J.-R.J., Kelly, J.J., Watson, D.C.D., Tessier, L.L., Lanthier, P.H.P., Jarrell, H.C.H., Cadotte, N.N., Michael, F.F.S., Aberg, E.E., et al. (2002). Structure of the *N*-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. J Biol Chem *277*, 42530–42539.

Zhang, G., Brokx, S., and Weiner, J. (2006). Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. Nat Biotechnol *24*, 100–104.

Zhang, Z., and Palzkill, T. (2003). Determinants of binding affinity and specificity for the interaction of TEM-1 and SME-1 beta-lactamase with beta-lactamase inhibitory protein. J Biol Chem *278*, 45706–45712.