ELUCIDATING A NOVEL MECHANISM OF DNA SILENCING CAUSED BY ENVELOPE STRESS 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

Ritsdeliz Pérez-Rodríguez

February 2010

ELUCIDATING A NOVEL MECHANISM OF DNA SILENCING CAUSED BY ENVELOPE STRESS IN *ESCHERICHIA COLI*

Ritsdeliz Pérez-Rodríguez, Ph. D.

Cornell University 2010

A widespread feature in the genomes of most bacteria and archaea is an array of clustered, regularly interspaced short palindromic repeats (CRISPRs) that, together with a group of CRISPR-associated (Cas) proteins, mediate immunity against infection by extrachromosomal agents such as phages. RNA molecules derived from the CRISPRs are predicted to guide the Cas proteins against foreign genetic elements with sequences matching the spacers. Here we found that bacteria can specifically silence plasmid DNA encoding an exported protein that induces envelope stress. Elimination of this plasmid DNA required base-complementarity between spacer DNA and the plasmid, and the participation of the Cas proteins. Of these, the CasC, D, and E proteins form a stable ternary complex and function as the catalytic core of the Cas system to process the pre-CRISPR RNA into the mature form. Additionally, CasE and several other Cas proteins possess DNase activity suggesting that these enzymes have a role in the effector function of the CRISPR-Cas system. We conclude that the Cas interference machinery can target DNA directly using CRISPR-derived small RNAs as guides and provides a general defense mechanism that is not limited to phage infection.

BIOGRAPHICAL SKETCH

Ritsdeliz Pérez-Rodríguez is the youngest of three girls born to Héctor Pérez and Rita Rodríguez in Caguas, Puerto Rico. Her interest and passion for Chemistry and Mathematics led her to an early acceptance in the Chemical Engineering Department at the University of Puerto Rico, Mayagüez Campus. During her BS studies she worked as a Process Engineer COOP student in Pharmacia, Arecibo, and also obtained the Engineering in Training (EIT) license. Immediately after completing her BS Degree in 2001, she began her Masters Degree in the same department under the guidance of Dr. José A. Colucci and graduated in 2004. Her MS thesis research was based on renewable energy, one of the most important topics of interest nowadays, by studying autothermal reforming for fuel cells in automotive applications. In 2004, she began her PhD studies in the Chemical and Biomolecular Engineering Department at Cornell University in Ithaca, NY, under the guidance of Dr. Matthew DeLisa. The main objective of her dissertation was to characterize and understand the existence of a possible RNAi machinery in prokaryotes. She will be receiving her Doctorate of Philosophy in January 2010.

To the ones I love and owe my life:

God for being my strength

My family for your unconditional support

Thank you

ACKNOWLEDGEMENTS

First of all, I would like to acknowledge God, without Him, my work would not have been as successful as it was, and this thesis would never have been written. Second, my special thanks to my advisor Matthew DeLisa for giving me the opportunity to do something different in my academic life. Changing from the traditional Chemical Engineering field to Biotechnology was a real challenge for me and added a positive spike of curiosity to those who read my *curriculum vitae*. Thanks for believing that I could do something completely new for everyone and for letting me follow my own instincts and pioneer my own research. My gratitude also goes to my committee members, Dr. Michael Shuler and Dr. Alan Collmer.

I would also like to aknowledge Cornell University for funding through a Sloan Fellowship and the NYSTAR James D. Watson Award (to M.P.D).

Thanks to the DeLisa Research Group for teaching me the basic molecular cloning techniques and strategies, a background that I didn't have when I came to the lab for the first time. Thanks to Rob for his crazy cloning strategies that always worked, and to Tom for suggesting more experiments that converted my papers in stronger versions. Specially, thanks to Charles who, with his microbiology expertise, also helped me in the understanding of basic molecular biology. Thanks for standing up for me at my oral presentations and for being a true friend outside the lab. I will miss you tons.

Also, my acknowledgements to Chris, you came in the right moment to the lab. Thanks for your input and showing me your special techniques for running the "sexiest gel ever"! Thanks to you, Charles, Mark, Tom and Rob for making fun of my Spanish accent, I laughed and learned a lot too! I will

never forget that. Thanks to Didi, you made my life so easy in the lab. Your happiness was very contagious. Thanks to Adam, Rob, Charles, Cassie, Jason, and Tom for kindly proofreading my thesis.

Thanks to my family, mom, dad, Michelle, Landy, Yasmin, Mariano, Derek and Megan; and to my best friends Luz and Alejandro, I have no words for describing how much you helped me throughout my PhD studies. Thanks for your support, understanding, and for never forgetting how much I love you and how important you are in my life. After all these years being apart from you guys, I am glad that it is time to go back home.

Thanks to my Puerto Rican, Colombian, Dominican, and all my Latin friends here at Cornell for being part of my life here in Ithaca. Thanks to you guys, I never forgot my roots and how proud I am of being Puerto Rican especially at the Latin dance parties. My special thanks to Alexis, Jeisa, and Karen, you guys were my truly and sincere support in my moments of loneliness. I love you and will never forget you. Thanks to Edward and Robin for taking care of me when I needed the most, you were like my brother and sister here in Ithaca. Finally, but not least, thanks to my friends at the Sloan Lunches, Anthony, Jeisa, Miguel, Carlos Ivan, and Gabriel for making me laugh even at my most stressful moments. I will miss you too.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The ability to respond to changes in the internal and external environment by modifying the metabolism or activity of genes is important for most living organisms. Bacteria need to employ envelope stress responses for a variety of processes including protein folding, cell wall biosynthesis, and pathogenesis (Raivio and Silhavy, 2001). Most of these responses are controlled by heat shock and extracytoplasmic function (ECF) sigma factors, or by two-component regulatory systems that are mainly comprised of a sensor kinase and a response regulator.

The best-characterized heat shock sigma factor is σ^{32} in *Escherichia coli*. The σ^{32} regulon encodes a set of chaperones and proteases that play an important role in the heat-shock response in *E. coli* (Yura, 1999). This response is a cellular protective reaction which allows cells to adapt to metabolic and environmental changes and to cope with stress-induced damage of intracellular proteins. Most of the heat-shock proteins (HSPs) are chaperones such as DnaK-DnaJ-GrpE and GroEL-GroES, or ATP-dependent proteases such as Lon and ClpAP (Arsene et al., 2000). The major roles of these HSPs include protein folding, assembly, transport and repair under stress conditions.

When *E.coli* cells are grown at 37°C and are suddenly exposed to 42°C, the cellular level of σ^{32} increases, thus enhancing the transcription of gene encoding HSPs (Yura, 1993). This process initiates the expression of the *rpoH* gene, which encodes σ^{32}. The resulting sigma initiator factor binds to the RNA polymerase (RNAP). These two proteins, forming the σ^{32}-RNA

complex, concomitantly activate the transcription of HSPs, where more than 20 HSPs are synthesized either to fold (chaperones) or to degrade (proteases) misfolded proteins (Bukau, 1993). Following this process is an adaptation period is followed, where the synthesis of these proteins decreases to reach a new steady-state level inside the cell which is mediated by the DnaK chaperone system (DnaK-DnaJ-GrpE). This system can inactivate or degrade σ^{32}, and repress *rpoH* translation during the "shut-off" phase of the heat shock response (Tatsuta et al., 1998).

Besides the DnaK chaperone system, there are several other heat-shock cognate homologues of DnaK and DnaJ chaperones that have been identified, although their exact function in heat shock regulation remains unknown (Vickery et al., 1997). However, it has been shown that CbpA, a DnaJ analog, has an overlapping regulatory function with DnaJ (Tatsuta et al., 1998). Also, it has been suggested that some small HSPs, such as IbpA and IbpB, cooperate or act together with other HSPs including DnaK, DnaJ, GroEL, HtpG, and ClpB (Thomas and Baneyx, 1998). These data suggest that the cells employ a multichaperone network where most general chaperones participate in modulating the heat shock response.

Other than exposure to extreme temperature, chemical stresses can also induce the heat shock response in bacteria. Exposure to various chemicals including ethanol, 2,4-dinitrophenol, sodium azide, and heavy metals, are known to induce the bacterial heat shock response (Van Dyk et al., 1994). Remarkably, the over production of HSPs is produced by inducing σ^{32} synthesis at the level of transcription rather than translation (Lopez-Sanchez et al., 1997).

The heat shock stress systems discussed above respond to perturbations of the microbial cell by acting inside the cytoplasm, but there are other systems that function outside of this compartment. Such responses are called extracytoplasmic stress responses (ESRs) and have been shown to play a critical role in the pathogenesis of Gram-negative bacterial pathogens such as *E. coli O157:H7*, and *Pseudomonas aeruginosa* (Rowley et al., 2006). These responses target stresses that affect cell envelope components including periplasmic and outer membrane proteins (OMPs) (Ruiz et al., 2006). The best characterized of these responses are those mediated by the ECF sigma factor, σ^E, and the two-component regulator CpxAR (Danese and Silhavy, 1997; De Las et al., 1997). Two additional ESRs include: phage-shock response and the recently discovered BaeSR-regulated response (Jones et al., 2003; Raffa and Raivio, 2002).

The haloenzyme product σ^E of the gene *rpoE*, is an essential sigma factor in some, but not all, bacteria (Craig et al., 2002; Humphreys et al., 1999). This ESR is mainly comprised of *rpoE* and the downstream genes *rseABC*. However, many other genes have been identified to be involved in this response and regulated σ^E. These genes encode proteins involved in folding or degradation of misfolded proteins in the periplasm, thiol:disulfide oxido-reductases (i.e., protein disulfide isomerase II, involved in disulfide-bond formation), those involved in lipopolysaccharide biogenesis or modification, and some others whose functionality is still not known (Alba and Gross, 2004; Duguay and Silhavy, 2004; Rezuchova et al., 2003).

The activation and regulation of σ^E is very similar to that of σ^{32}. Briefly, σ^E is inactive when bound to its cognate membrane-bound anti sigma-factor, RseA. When the membrane of the cell is perturbed, σ^E-RseA complex will

disassociate and σ^E will bind to the core of RNAP. Finally, the resulting complex σ^E-RNAP will activate the σ^E-dependent promoters that would lie upstream of the *rpoE rseABC* operon (Miticka et al., 2003). Like σ^{32}, the increased transcription of this operon will lead to very high levels of σ^E and RseA. Consequently, the regulator RseA will switch off the σ^E response once the activating stimuli are removed.

Some stress conditions that activate the σ^E pathway in *E. coli* are heat shock (σ^E-RNAP can activate the transcription of *rpoH* temperatures of 50°C or more) (Yura, 1999), cold shock, ethanol exposure, carbon starvation and oxidative stress (Rowley et al., 2006). All of these conditions lead to the accumulation of outer membrane proteins in the periplasm. In a proposed and novel mechanism, this accumulation would be the signal for σ^E activation through the protease DegS. This mechanism is termed regulated intra-membrane proteolysis (RIP), used by both prokaryotes and eukaryotes. During RIP, DegS will cleave the membrane-bound RseA to release σ^E in the cytoplasm and begin the transcription of the *rpoE rseABC* operon (Ehrmann and Clausen, 2004).

Aside from regulating bacterial functions during environmental stress conditions, σ^E is also involved in virulence pathways during pathogenesis (Rowley et al., 2006). The precise role of this sigma factor in some pathogenic bacteria is still not well defined, but it is thought to be complex, and likely varies among pathogens. Nonetheless, it is reasonable to believe that the mechanism of *P. aeruginosa* is similar as described above because σ^E in *E. coli* is homologous and functionally equivalent to the AlgU in *P. aeruginosa* (Govan et al., 1992). Moreover, the organization of the *algU* operon *algU*

mucA mucB mucC is equivalent to that in *E. coli rpoE rseA rseB rseC* (Kunst et al., 1988).

Another system that is generally used by bacterial cells for sensing and responding to extracytoplasmic conditions are the two-component signal transduction systems. Such systems consist specifically of a transmembrane sensor kinase and a cytoplasmic response regulator. The best described two-component system in responding to envelope stress in *E. coli* is the CpxAR or Cpx system (Pogliano et al., 1997). This system can control over 100 proteins in response to envelope stresses, some of them overlapping with the σ^E regulon (Raivio and Silhavy, 1999). CpxR, CpxA, and CpxP proteins are members of the Cpx regulon and act as a response regulator, a sensor kinase, and as the negative-feedback regulator of the system, respectively. A model for the regulation of Cpx system has been proposed where, first, this system is triggered by envelope stresses and, second CpxP disassociates from CpxA, activating the CpxA kinase and for the phosphorylation of CpxR (Pogliano et al., 1997). Finally, phosphorylated CpxR regulates the gene expression by binding to the specific consensus sequence 5′-GTAAA-N5-GTAAA-3′ positioned upstream of the Cpx regulon members (De Wulf et al., 2002).

Like σ^E, the Cpx regulon is also associated with the virulence of some pathogenic bacteria. In *E. coli*, Cpx regulates pili and fimbrial assembly genes, including *dsbA* and *ppiA*. Enteropathogenic *E. coli* without *cpxR* produce less type IV bundle-forming pili and show attenuated adherence to eukaryotic cells (Nevesinjac and Raivio, 2005). Thus, bacterial regulators of envelope stress often have critical roles in the virulence of pathogens.

A third two-component system that has been identified as an envelope stress signal transduction pathway in *E. coli* is the BaeSR system. During

certain conditions of envelope stresses, such as the formation of spheroplasts, the presence of indole (a toxic compound that disrupts the cell wall), or the expression of misfolded pilus protein PapG, the BaeSR signal transduction pathway can mediate adaptation to such stresses. This is done by increasing expression of *spy*, a known member of the Cpx regulon (Raffa and Raivio, 2002). Interestingly, both BaeSR and Cpx induce high levels of *spy* expression, but they respond differently to environmental signals. For example, both pathways induce *spy* expression during spheroplast formation and over expression of PapG, but the BaeSR pathway responds more dramatically to the presence of indole. Besides controlling the expression of *spy* in response to envelope stresses, BaeSR does not affect expression of other known regulated genes from the Cpx regulon. Nonetheless, double knockouts of both BaeR and CpxR result in improved sensitivity to envelope stresses relative to either single knockout alone. For this reason, Raffa and Raivio have proposed that the BaeSR two-component system is distinct from the Cpx regulon and it is a third envelope stress response in *E. coli* that induces expression of a diverse set of adaptive genes (Raffa and Raivio, 2002).

A fourth ESR system named the phage shock response is also proposed in *E. coli.* Here, the phage shock protein (PSP) is induced following infection with filamentous phage (Brissette et al., 1990). The PSP response is mediated by the *pspABCDE* operon, where the first gene *pspA* can also be induced by several environmental stress situations such as high temperature, osmotic stress, and ethanol/organic-solvent exposure (Weiner and Model, 1994). The PspA protein is believed to function in the maintenance of the proton-motive force (PMF) across the inner membrane (Kleerebezem et al.,

1996; Model et al., 1997). Moreover, PSP can compensate for the absence of σ^E in *Salmonella typhimurium rpoE* mutants in stationary phase (Becker et al., 2005). This is another demonstration of how different ESR pathways interact with each other at distinct envelope stress conditions. The role of the PSP regulon in bacterial virulence is not well known yet (Rowley et al., 2006).

The involvement of ESRs in survival or in bacterial pathogenesis has only been elucidated in a few bacterial species, and there is still much to learn about it. This includes determining the role of each ESR regulon and the functionality of each one of its members. Despite the lack of full understanding of this regulon, it is known that its members can be exploited for the development of new vaccines and chemotherapeutic agents. For instance, the mutation of *Salmonella enterica* HtrA protease, which is important for the virulence of many Gram-positive and Gram-negative bacterial pathogens, will make a safer organism for effective live oral vaccines (Loosmore et al., 1998). In addition, the members of the ESR regulon could be clinically relevant because they are potential targets for novel antimicrobial strategies. An example of this was shown by Yang and coworkers. By inhibiting several ESR pathways of *E. coli*, they could reduce the bacterial resistance to novel antibiotics (Yang et al., 2004). Specifically, an increase in the Minimum Inhibitory Concentration (MIC) of several novel antibiotic agents was observed with the over expression of σ^E in *E. coli*.

Besides being stimulated at the DNA and protein level, the cytoplasmic and extracytoplasmic envelop stress responses, could also be induced at the RNA level (Waters and Storz, 2009). Bacteria use heterogeneous group of molecules, namely RNA regulators, to respond to a broad range of physical responses. RNA regulators use various mechanisms to modulate these

responses and one of the best characterized classes is called riboswitches. These RNA sequences fold into structures amenable to conformational changes upon binding of small molecules. The riboswitches sense and respond to the envelope stress caused by different environmental conditions of the cell. Another class of regulators that respond to envelop stress is called small RNAs (sRNAs). These small transcripts are the largest and most studied regulators and are known to bind to proteins and antagonize their regulatory functions. They also base pair with other RNAs and modulate the translation of mRNAs. Lastly, another class of cell regulators is the recently discovered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNAs that act together with their associated Cas (CRISPR-associated) genes to react to different envelope stresses. These CRISPR RNAs contain short homology regions to phages and plasmid DNA sequences. This homology is believed to be the cause of the bacterial interference with phage infection and plasmid conjugation by a mechanism that is still not well understood (Makarova et al., 2006; Marraffini and Sontheimer, 2008).

Riboswitches bind to small molecules and adopt different conformations in the presence or absence of metabolite. In a very particular way, they directly regulate the genes related to the uptake and use of metabolites (Montange and Batey, 2008). A broad diversity of riboswitches has been identified in bacteria. Surprisingly, as many as 2% of *Bacillus subtilis* genes are known to be regulated by riboswitches that would bind to some metabolites, such as thiamin pyrophosphate, lysine and guanine (Waters and Storz, 2009). Interestingly, researches have been able to identify the

physiological role of a gene product by taking advantage of the presence of a riboswitch upstream of such gene.

Riboswitches consist of two main regions: the aptamer and the platform region (Nudler and Mironov, 2004). The aptamer will bind to a ligand (usually a metabolite), and the platform will regulate gene expression by acquiring specific RNA conformations. This mechanism starts with the binding of the aptamer to the metabolite ligand and ends with changes in conformation of the riboswitches. These conformations are usually hairpin structures that form or disrupt transcriptional terminators or antiterminators, inhibiting or allowing gene expression by occluding or exposing ribosome-binding sites. In general, riboswitches in Gram-negative bacteria inhibit translation, whereas riboswitches in Gram-positive bacteria affect transcriptional attenuation (Nudler and Mironov, 2004). Bacterial genomes generally have large, tightly clustered biosynthetic operons. Therefore, it is speculated that more resources would be wasted if, in the absence of riboswitches, the full-length transcripts of these regulons are synthesized. In this manner, riboswitches allow for more parsimonious gene expression in bacteria.

In addition to riboswitches, there are certain sRNAs that respond to envelope stresses. However, sRNAs act by modulating protein activity in bacterial cells. Protein-binding sRNAs such as CsrB and 6S antagonize the activity of their cognate proteins by acquiring the same structures of other nucleic acids. For instance, *E. coli* CsrB and CsrC RNAs modulate the activity of CsrA, which is a protein that regulates carbon usage and mobility of cells upon entry in stationary phase, starvation or nutrient-poor conditions (Babitzke and Romeo, 2007). It has been shown that CsrA protein binds to GGA hairpins in the 5'UTR of the target mRNAs altering and affecting the stability of

mRNA and gene expression. When nutrients are depleted in the environment, *csrB* and *csrC* are induced by the BarA-UvrB two component regulators by an unknown mechanism. CsrB and CsrC RNAs each have numerous GGA binding sites for CsrA, so if the sRNA levels of CsrB and CsrC increase, CsrA will be sequestered by these sRNAs away from other leader mRNAs, thereby, preventing translation. Like CsrB and CsrC, the protein-binding sRNA called 6S also has an antagonizing role; when cellular 6S sRNA levels increase, they titrate the σ^{70}-RNAP complex away from DNA promoters, inhibiting transcription of genes (Trotochaud and Wassarman, 2005).

In general, sRNAs can be classified in two classes: (i) those that have extensive potential for a 100% base pairing with the target RNA, called *cis*, and (ii) those sharing only a limited complementarity, called *trans* (Waters and Storz, 2009). Most of *cis*-encoded sRNAs reside on mobile genetic elements such as plasmids or transposons. *Cis*-sRNAs are responsible for maintaining the appropriate copy number of mobile elements, thereby preventing high gene expression levels which may burden the cell (Brantl, 2007; Wagner et al., 2002). This may occur in a number of proposed mechanisms. One mechanism, for instance, involves the inhibition of replication primer formation (ColE RNA I) and transposase translation (Tn10 pOUT RNA) (Waters and Storz, 2009).

In addition to extrachromosomal *cis*-sRNAs, some *cis*-sRNAs are located on the bacterial chromosomes; however, their physiological roles are poorly understood. Some chromosomal *cis*-sRNAs can degrade and/or repress translation of target mRNAs, such as proteins that are toxic at high levels (Fozo et al., 2008). This mechanism of degradation occurs when the *cis*-sRNAs and the target mRNA share almost a 100% homology, or extended

regions of complementarity (often 75 nucleotides or more) (Brantl, 2007; Wagner et al., 2002). A good example is the *E. coli cis*-sRNA called IstR. It is located next to *tisAB*, a gene that encodes the toxic protein TisB (Vogel et al., 2004). Although IstR is not a true antisense RNA (it does not share 100% homology with the toxin mRNA), it can target TisB for degradation by sharing regions of 19 and 23 nucleotides of perfect complementarity. By this means, IstR *cis*-sRNA prevents TisB translation during normal growth and could also limit the toxicity induced under stress conditions. The latter occurs during the SOS response to DNA damage (Kawano et al., 2007). It has been proposed that this toxin-antitoxin module could slow cell growth under stress conditions to give the cells time to repair DNA damage or adjust to their new environment. Additionally, other modules conserved in bacterial chromosomes can act as a defense mechanism against plasmids bearing homologous portions of the target toxin mRNA. This will happen only if the chromosomal *cis*-antisense sRNA can repress the expression of the protein in the plasmid by base pairing to the mRNA. These chromosomal versions of sRNAs and their mechanisms of action are not yet clearly defined.

In contrast to the *cis*-encoded antisense sRNAs, the *trans*-encoded sRNAs share only limited complementarity with the target mRNA and are shown to be analogous to eukaryotic sRNAs (Gottesman, 2005). But like the *cis*-encoded sRNAs, *trans*-encoded sRNAs mechanism usually involves the base pairing between the sRNA and the target mRNA, leading to the repression of protein levels through inhibition of translation or mRNA degradation. Interestingly, these sRNAs can also activate the expression of their target mRNAs through a mechanism whereby the inhibitory secondary structure of the mRNA that normally sequesters the ribosome binding site will

be disrupted by base pairing of the sRNA (Prevost et al., 2007). Therefore, base pairing between a *trans*-encoded sRNA and its target can promote both transcription termination and activation.

The *trans*-encoded sRNAs usually base pair with multiple mRNAs (Gottesman, 2005). These sRNAs make very limited contact with the target mRNAs rather than extended stretches of 100% complementarity. Here, only 10-25 nucleotides participate in the base pairing of sRNAs and target mRNAs, and only a fraction of these nucleotides can be critical for regulation (Waters and Storz, 2009). For this reason, the presence of the RNA chaperone Hfq is necessary because it will facilitate the RNA-RNA interactions due to this limited complementarity mentioned before (Aiba, 2007). Also, this chaperone allows sRNAs and target mRNAs to "sample" potential complementarity and increase the concentrations of sRNAs and mRNAs (Waters and Storz, 2009). Hfq chaperone is critical in RNA-RNA interactions because its absence affects the regulation of target mRNAs in *E. coli.* Furthermore, multiple sRNAs coimmunoprecipitate with Hfq, making this technique useful in identifying novel *trans*-encoded sRNAs (Zhang et al., 2003).

Beyond being a facilitator for base pairing, Hfq also prevents sRNAs from degradation. Researchers have shown that in *hfq* mutant strains, *E. coli* sRNAs are not stable, presumably because the base pairing with mRNAs is completely absent (Brennan and Link, 2007). Finally, it is worth mentioning that, in contrast with *cis*-sRNAs, most of these *trans*-encoded sRNAs are synthesized under very specific growth and stress conditions. For instance, in *E. coli*, *trans*-RNAs are induced mainly by (i) low iron (Fur-repressed RyhB), (ii) oxidative stress (OxyR-activated OxyS), (iii) outer membrane stress (σ^E-induced MicA and RybB), and (iv) changes in glucose concentration (CRP-

repressed Spot42 and CRP-activated CyaR (De Lay and Gottesman, 2009). The fact that the base pairing of *trans*-sRNAs is limited and regulates multiple mRNA targets offers the possiblity of the existence of a single global sRNA modulating a particular physiological response at the post transcriptional level (Bejerano-Sagie and Xavier, 2007). It is possible that every major transcription factor controls the expression of one or more sRNAs (Waters and Storz, 2009).

Recently discovered regulatory RNAs with some similarities to the eukaryotic siRNAs that lead to gene silencing are called CRISPR RNAs. In bacteria, these CRISPR RNAs supply resistance to bacteriophages (Sorek et al., 2008) and prevent plasmid conjugation (Marraffini and Sontheimer, 2008). CRISPRs are found in almost 40% of sequenced bacteria and in almost 90% of sequenced archaea (Sorek et al., 2008). These sequences vary among species, but are the same within given species. These sequences can range from 24-47 base pairs (bp), where 5 to 7 bp are palindromic. CRISPRs can be repeated in the genome 2 to 249 times, and also have an AT-rich leader sequence located upstream of their region, which is about 300-500 bp long (Jansen et al., 2002; Mojica et al., 2000). These leader sequences are almost identical among some related species (~80%), and are found nowhere else in the genome of bacteria. Remarkably, these sequences are not found in *E. coli* and *Sulfolobus sulfataricus* regardless of their multiple CRISPR-spacer loci (Jansen et al., 2002). Also, these repeats are separated by unique spacers that have variable lengths ranging from 26 to 73 bp, and are found to be homologous to the DNA of phages and plasmids (Makarova et al., 2006).

Adjacent to the CRISPR-spacer region, there are two to six *cas* genes related to most CRISPR systems in bacteria. Interestingly, these *cas* genes

are never found in bacterial genomes lacking the CRISPR-spacers region. The exact functions of all Cas proteins are not well understood yet, but bioinformatic studies show that they contain RNA- or DNA- binding domains, helicase motifs, and endo- or exonuclease domains (Makarova et al., 2006). Also, based on these studies, the *cas* genes encode proteins with functionalities analogous to the eukaryotic RNA interference enzymes.

CRISPRs-spacers, together with the *cas* genes, can provide resistance against phages and plasmids by homology of the spacers to the target mRNA or DNA (Marraffini and Sontheimer, 2008; Sorek et al., 2008). The entire region of CRISPR-spacers DNA in the bacterial genome is transcribed into large RNAs that will be sequentially cut into small RNAs by the Cas genes. The resulting small RNAs will be comprised of a single CRISPR-spacer-CRISPR unit. Cas members such as CasA, B, C, D, and E, can associate to form a complex (Cascade complex) as part of the antiviral defense (Brouns et al., 2008). Specifically, CasE is responsible for the processing of the full length CRISPR RNA transcript (our own observations).

Interestingly, it has been determined that the spacers corresponding to some portions of the phage DNA are incorporated by a mysterious mechanism into the bacterial chromosome. These spacers confer resistance to the cells to survive subsequent infections either with the same phage or others having the same sequence (Barrangou et al., 2007). Spacers are acquired at the beginning of the CRISPR-RNA array as new infections occur. Thus, the first spacers encountered in this array (5'-) represent a DNA portion of the phage that infected the cells most recently. Similarly, the last spacers in this array (-3') represent DNA portions of more ancient phage infections. Experimental data show that, when synonymous mutations are made to the spacers or in

the phage genome, the resistance of the cells is abolished. Moreover, when phage DNA portions are engineered as spacers into the bacterial CRISPR-spacer array, the resistance is acquired again, thus providing *de novo* immunity to bacteria against the phage (Barrangou et al., 2007). These observations were similar to those made in a recent set of data where spacers were found to correspond to parts of invading conjugative plasmids in *Staphylococcus epidermidis* (Marraffini and Sontheimer, 2008).

Taking the bioinformatic analysis made by Makarova and co-workers (Makarova et al., 2006) together with these experimental observations described above, a model for CRISPR RNA activity has been proposed. Briefly, the CRISPR DNA array is transcribed into a long RNA bearing hairpin secondary structures which are created by the palindromes present in the CRISPR repeats. Such hairpins are recognized by the Cascade complex and processed into small RNAs that consist of a CRISPR-Spacer-CRISPR unit (Brouns et al., 2008). The single stranded small RNAs are thought to be then retained and directed by some Cas proteins to a homologous sequence of phage or plasmid for *cis*- or *trans*- base pairing and targeting mRNA or DNA (Barrangou et al., 2007; Brouns et al., 2008).

After the discovery of a possible RNAi system in bacteria, it was believed that RNA was the target for cleavage and degradation of invasive DNA. But, it has recently been shown that DNA is the target that will lead to such degradation. Researchers have shown that an insertion of an intron into the target gene DNA in a conjugative plasmid abolished gene silencing by CRISPRs RNAs, even though the uninterrupted target sequence will eventually be regenerated in the spliced mRNA (Marraffini and Sontheimer, 2008). This result, amongst other similar ones, provides the first experimental

evidence for another role of the CRISPR interference. That is, the CRISPRs RNAs can also target DNA directly before transcription. The exact mechanism involved in this degradation is unknown. Further studies will provide a better understanding of the similarities and differences to the eukaryotic RNA interference system and may also provide useful tools for researchers to genotype strains and to study some other bacterial phenomena such as horizontal gene transfer and microevolution (Waters and Storz, 2009).

In general, riboswitches, protein-binding sRNAs, and *cis*- and *trans*-encoded base pairing sRNAs, mediate responses to changes in environmental conditions by modulating metabolic pathways and/or stress responses in bacteria (Waters and Storz, 2009). Riboswitches regulate biosynthetic genes by directly sensing variations in the concentration of metabolites, while the protein-binding CsrB and 6S sRNAs control the expression of genes in response to starvation by repressing the activity of global regulators. Likewise, the contribution of *cis*- and *trans*-encoded sRNAs may also be involved in cell survival. These sRNAs could alter the metabolism inside the cells and modulate the translation or repression of the synthesis of unneeded proteins in response to various cell life-threatening stresses. Lastly, the specific role of the CRISPR RNAs regulators in response to stress conditions in bacteria is still unclear. However, we have discovered that envelope stress likely caused by the failed attempt of *E. coli* to transport folded proteins through the inner membrane can activate interference by means of CRISPR RNAs. In the following chapters, the complete story of how we made this discovery is fully depicted.

CHAPTER 2

AN ESSENTIAL ROLE FOR THE DNAK MOLECULAR CHAPERONE IN STABILIZING OVER-EXPRESSED SUBSTRATE PROTEINS OF THE BACTERIAL TWIN-ARGININE TRANSLOCATION PATHWAY[1]

1. Introduction

In bacteria, most proteins with extra-cytoplasmic function are translocated across the inner cytoplasmic membrane by either the general secretory (Sec) pathway or the twin-arginine translocation (Tat) pathway, both of which recognize their substrates in part by cleavable N-terminal signal peptides. In the case of the Sec pathway, the translocation pore formed by SecYEG is a narrow (5-8Å) conduit that precludes passage of unfolded proteins (Van den Berg et al., 2004). Thus, proteins that transit the Sec pathway must be maintained in a largely extended conformation, a feat that is accomplished with assistance from the cytoplasmic chaperone SecB (Collier et al., 1988; Driessen et al., 2001; Knoblauch et al., 1999; Kumamoto, 1989; Kumamoto and Beckwith, 1983; Lecker et al., 1990; Liu et al., 1989; Watanabe and Blobel, 1989) and the general molecular chaperones DnaK-DnaJ and GroEL-GroES (Phillips and Silhavy, 1990; Wild et al., 1992; Wild et al., 1996). Whereas the Sec system processes proteins that are unfolded, the hallmark of the bacterial Tat machinery is the ability to post-translationally translocate proteins that have undergone folding in the cytoplasm, prior to export (Berks et al., 2000; Robinson and Bolhuis, 2001). One of the most

[1] Adapted with permission from Pérez-Rodríguez, R, Fisher AC, Perlmutter, JD, Hicks MG, Chanal, A., Santini, CL, Wu, LF, Palmer, T. and DeLisa, MP (2007). "An Essential Role for the DnaK Molecular Chaperone in Stabilizing Over-expressed Substrate Proteins of the Bacterial Twin-arginine Translocation Pathway" in Journal of Molecular Biology.

remarkable features of the bacterial Tat system is its ability to selectively transport only those substrates that are correctly folded and assembled prior to transport (DeLisa et al., 2003; Fisher et al., 2006; Kim et al., 2005; Rodrigue et al., 1999; Sanders et al., 2001), leading to speculation that a folding quality control mechanism governs Tat transport.

In light of such a quality control system, a two-tiered mechanism to prevent premature export of incompletely folded and/or assembled substrates was proposed (Turner et al., 2004) in which the late stages of substrate folding are "sensed" by the Tat machinery directly while earlier stages of folding are proofread by cytoplasmic chaperones. For instance, a large proportion of Tat substrate proteins acquire complex redox-active multi-atom cofactors (e.g., Fe-S clusters, molybdopterin) in the cytoplasm and require correct assembly before transport can proceed (Berks, 1996). How such complex cofactor-containing proteins are prevented from engaging the translocation machinery prior to cofactor attachment has been partially resolved by the discovery of Tat system-specific chaperones such as redox enzyme maturation proteins (REMPs) (Dubini and Sargent, 2003; Jack et al., 2004; Turner et al., 2004). REMPs are cytoplasmic chaperones that are involved in the maturation of redox enzymes. A subset of REMPs are postulated to bind and sequester Tat signal peptides until their cognate Tat-targeted redox enzyme is fully folded and assembled and competent for transport, a mechanism known as the "shelter" model (Berks et al., 2000). The best-characterized REMP is *E. coli* DmsD which has been shown to bind to the Tat signal peptides of DmsA (Ray et al., 2003) and TorA (Oresnik et al., 2001), and plays an important role in DmsA maturation and localization (Berks et al., 2003; Ray et al., 2003). Numerous other REMPs have been identified such as the TorD protein which

pairs with TorA (Jack et al., 2004) and it has recently been suggested that all Tat substrates may possess a cognate leader-binding chaperone (Palmer et al., 2005).

It is noteworthy however that not all Tat substrates attach complex cofactors. Notable examples include *E. coli* AmiA, AmiC, CueO and SufI (Berks et al., 2005) as well as numerous Tat-competent heterologous substrates such as alkaline phosphatase (DeLisa et al., 2003; Tullman-Ercek et al., 2006), β-lactamase (Fisher et al., 2006), green fluorescent protein (DeLisa et al., 2002; Santini et al., 2001; Thomas et al., 2001) and maltose binding protein (Blaudeck et al., 2003; Tullman-Ercek et al., 2006). Without the insertion of such cofactors and/or a corresponding REMP to monitor the progression of substrate maturation, it is unclear how the folding and transport of heterologous proteins lacking cofactors is coordinated so as to prevent membrane targeting of incompetent substrates. In these instances, it is envisioned that general folding catalysts may participate in the folding and proofreading of Tat substrates by sequestering misfolded proteins from the translocon until correct folding or proteolytic degradation has occurred (Fisher and DeLisa, 2004). Several lines of evidence implicate a role for general molecular chaperones in bacterial Tat transport including: (1) the chaperone trigger factor (TF) binds extensively to the TorA and SufI signal peptides although depletion or overexpression of TF had little effect on the kinetics and efficiency of the Tat export process (Jong et al., 2004); (2) GroEL was found to affect Tat-specific hydrogenase-1 maturation and assembly (Rodrigue et al., 1996) and was also found to associate with the Tat-specific amidase AmiA (Kerner et al., 2005); (3) DnaK specifically binds to the Tat leader peptide of DmsA (Oresnik et al., 2001); (4) members of the chaperone cascade including

TF, DnaK-DnaJ-GrpE and GroEL interact with the Tat-specific REMP DmsD (Howell and Turner, 2006); and finally, (5) overexpression of DnaK-DnaJ-GrpE, GroEL-GroES and TF has been shown to enhance the transport the Tat substrate preproPAC (Xu et al., 2005).

In addition to the direct involvement of cytoplasmic chaperones, a number of indirect observations suggest that factors other than TatABC are involved in Tat transport. For instance, two separate groups have reported that translocation of the Tat substrate SufI could not be observed in inside-out inner membrane vesicles (INVs) prepared from wt *E. coli* cells (Alami et al., 2002; Yahr and Wickner, 2001). Only in membrane vesicles that had been prepared from cells overexpressing TatABC from a strong T7 promoter could translocation be detected, albeit at a relatively low yield (at most 20%). Along similar lines, a number of native Tat substrates expressed *in vivo* have been observed to transit the inner membrane with efficiencies well below 100% (Chanal et al., 2003; Gon et al., 2000; Mikhaleva et al., 1999). Collectively, these observations suggest that cytoplasmic factors involved in substrate proofreading and/or targeting (akin to SecB) might help maintain export competence and enhance translocation efficiency *in vivo.* To test this hypothesis, we have systematically evaluated the effect of 21 well-characterized *E. coli* mutant strains for their ability to translocate GFP via the Tat pathway. Overall, our results indicate that the molecular chaperones DnaK-DnaJ are essential to maintaining the stability of Tat substrates in the cytoplasm prior to export. Consistent with this role, we observed that DnaK associates *in vivo* with at least one Tat-targeted substrate, a truncated variant of the Tat-specific enzyme trimethylamine *N*-oxide reductase (TorA). In addition, DnaK-DnaJ are able to promote efficient transport under conditions

where the Tat machinery is saturated by high levels of substrate, a finding that is expected to have important consequences for the use of the Tat system in biotechnology applications.

2. Results

2.1. Screen for cellular factors affecting expression of GFP via the Tat pathway reveals DnaK-DnaJ-GrpE and GroEL-GroES chaperone systems

In a previous search for factors that, when supplied *in trans*, confer enhanced Tat export of a short-lived version of the green fluorescent protein (hereafter GFP-SsrA), we identified the phage shock protein PspA as well as several other cytoplasmic proteins including the small heat shock chaperones IbpAB (DeLisa et al., 2004). While the above experiments did not reveal a physiological role for the cytoplasmic chaperones IbpAB in Tat transport (data not shown), they led us to hypothesize that other cellular factors may be involved in the Tat transport process. To test this, we employed a genetic screen to examine a collection of mutant *E. coli* strains for their ability to transport GFP through the Tat system. For these studies, we made use of a genetic reporter plasmid, pTGS, encoding a tripartite fusion between the *E. coli* TorA signal peptide and GFP-SsrA (TGS) that is transported specifically to the periplasm by the Tat pathway (DeLisa et al., 2002). Owing to the C-terminal SsrA degradation tag, this fusion protein is rapidly degraded in the cytoplasm by the proteases ClpAP and ClpXP (Farrell et al., 2005). Thus, the only cell fluorescence that emanates from cells expressing TGS arises from GFP that has been successfully transported to the periplasm prior to degradation in the cytoplasm ($t_{1/2}$ for GFP-SsrA in the cytoplasm is less than

1min (Farrell et al., 2005)). Since the TGS reporter is coupled to Clp-dependent degradation by virtue of its C-terminal SsrA tag, (DeLisa et al., 2002) we also employed a long-lived version of GFP carrying an N-terminal ssTorA signal peptide (TG) that was not linked to the SsrA proteolytic system to ensure that our results using TGS were not artifacts of this degradation mechanism.

To determine the effect of various cellular factors on TGS and TG expression and localization, a collection of well-characterized mutant strains and their isogenic parents were transformed with the Tat-specific reporter plasmid pMMB-TGS, an improved version of pTGS that expresses TGS from the IPTG-inducible vector pMMB (Bronstein et al., 2005) and, in parallel, with the pTG vector that expresses ssTorA fused to GFP from the arabinose-inducible pBAD33 (DeLisa et al., 2002). Induced cells were introduced into a flow cytometer and the mean cell fluorescence (MF) emitted from the wild-type parental strain was normalized by the MF of the corresponding isogenic mutant strain (Table 1). By this convention, an MF ratio of ~1 indicates no difference in Tat transport between the parent and the mutant strain, whereas an MF ratio >1 indicates that transport is reduced in the mutant relative to the parent and a value <1 indicates that transport is increased in the mutant relative to the parent. Results using these two expression plasmids were used in combination to rank order the relative contribution of each factor with respect to the transport of GFP and GFP-SsrA (Table 1). In general, the highest rank was given to MF ratios of ≥2 for both TGS and TG constructs, followed by MF ratios of ≥2 for only TGS, then MF ratios of ≥2 for only TG[2]

[2] Factors affecting TGS only were ranked higher than those affecting TG only, since TGS fluorescence is indicative of GFP that accumulates in the periplasm only due to Tat transport,

and finally MF ratios ≤1, indicating no effect on either TGS or TG under the conditions tested. Inspection of Table 1 reveals that the dnaK103, dnaK756, *ΔdnaKdnaJ*, groEL140, grpE280 and dnaJ259 mutant strains grown at permissive temperatures exhibited a significant fluorescence defect (≥2-fold MF ratio) following expression of TGS and TG. For all temperature sensitive mutants (e.g. dnaK103, dnaK756, groEL140, etc.), nearly identical defects in GFP expression and transport were observed when cultures were grown at 30 °C or 37 °C and shifted to the non-permissive temperature (44 °C, data not shown). Interestingly, even though the DnaK-DnaJ-GrpE triad is known to function as co-chaperones, the effect of dnaJ259 and grpE280 mutations on TGS and TG was far less dramatic relative to the MF ratios measured for dnaK103, dnaK756 and *ΔdnaKdnaJ* mutants. Also interesting was the observation that groEL140 imparted a large effect on both TGS and TG, but the groEL673 and groES30 mutations were found to affect only TGS and not TG; groES619 cells were actually more fluorescent than their parent strain. Taken together, mutations to the DnaK-DnaJ-GrpE chaperone system imparted the most substantial and consistent effect on Tat-targeted GFP-SsrA and GFP, and we focused our follow-up experiments on these chaperones.

---

whereas TG accumulates in both the cytoplasm and periplasm such that the MF ratio reflects differences in Tat substrate expression and transport between wild-type and mutant strains.

Table 1. *Cellular factors rank-ordered according to their effect on Tat transport of ssTorA-GFP-SsrA (TGS) and ssTorA-GFP (TG)*

Strain	pMMB-TGS	pTG
dnaK103	33.0 (145.4/4.4)	406.2 (1665.5/4.1)
dnaK756	27.0 (145.4/2.9)	289.2 (1665.5/3.3)
Δ*dnaKdnaJ*::kan	13.7 (95.6/7.8)	599.5 (1762.42/2.9)
groEL140[a]	26.0 (147.8/5.5)	200.8 (1446.0/7.2)
grpE280	4.4 (150.7/34.2)	2.2 (1201.5/542.9)
dnaJ259	2.0 (149.5/74.5)	4.6 (1665.5/363.4)
dmsD::cm[b]	25.4 (170.2/6.7)	1.3 (1840.4/1437.8)**
groEL673[a]	23.1 (147.8/6.4)	1.5 (1446.0/963.6)
groES30[a]	21.1 (147.8/7.0)	1.1 (1446.0/1354.8)
Δ*htpG*	7.0 (45.7/6.5)	1.1 (1596.3/1459.0)
Δ*clpB*::kan	3.4 (68.5/20.2)	1.6 (1596.3/982.2)
ftsH1(Ts)[c]	2.6 (66.2/25.4)	0.2 (116.3/650.0)
ftsH3::kan[c]	1.9 (338.0/179.9)	1.3 (1215.5/943.1)
lon::Tn*10tet*	1.3 (68.5/54.8)	2.3 (1732.0/756.3)
Δ*clpA*::kan	1.0 (68.5/65.4)	372.1 (1596.3/4.3)
ffh-77[d]	1.0 (13.8/14.3)	7.5 (6066.2/814.1)
dnaJ::Tn*10–42*[e]	1.0 (7.8/8.2)	2.6 (6200.5/2358.2)
dnaJ::Tn*10–42* Δ*cbpA*::kan[e]	1.0 (7.8/8.1)	3.7 (6200.5/1682.3)
Δ*cbpA*::kan[e]	0.9 (7.8/9.1)	1.0 (6200.5/6104.1)
Δ*torD*	1.1 (43.0/40.3)	1.0 (1775.8/1733.5)
secB::Tn*5*	0.9 (45.7/49.6)	1.2 (1459.0/1245.4)
degP41::kan[f]	0.9 (43.8/49.0)	1.0 (845.5/884.1)
ibp1::kan	0.8 (68.5/81.2)	1.0 (1733.0/1810.6)
tig::cm	0.8 (28.6/36.2)	1.0 (1998.9/1911.0)
groES619[a]	0.5 (147.8/328.4)	0.6 (1446.0/2320.3)

Data are reported as the ratio of mean fluorescence (MF) measured for wt cells relative to MF for mutant cells measured (*MF wt/MF mutant*). Data were taken 4–6 h post induction with 0.1 mM IPTG induction of pMMB-TGS and 0.2% arabinose induction of pTG. MF=mean fluorescence. All fluorescence activity values reported are the averages of three replicate experiments (standard error, <0.5%). All strains are derived from MC4100 unless otherwise noted. Plasmid pTrc-TG was used in place of pTG due to an antibiotic-resistance conflict.

[a] parent, B178.
[b] parent, TG1.
[c] parent, AR3289.
[d] parent, HPT57.
[e] parent, W3110.
[f] parent, KS272.

2.2. Instability of GFP and MBP in cells devoid of a functional DnaK protein is Tat specific

Following induction of pMMB-TGS or pTG, *ΔdnaKdnaJ* cells produced no measurable cell fluorescence above the background of cells lacking a GFP construct, whereas wild-type cells expressing the TGS and TG constructs were highly fluorescent (Figure 1a and b). To ensure that this lack of

fluorescence was not due to a general GFP folding defect in the *ΔdnaKdnaJ* strain background, we expressed GFP without a signal peptide in the cytoplasm of *ΔdnaKdnaJ* cells and observed strong cell fluorescence relative to wild-type cells expressing the same cytoplasmic GFP construct (Figure 1(c)). To determine if the non-fluorescent phenotype was attributable to DnaK or its co-chaperone DnaJ individually, we expressed TGS and TG in dnaK756 cells and, separately, in dnaJ259 mutant cells. Interestingly, whereas dnaK756 cells exhibited the same nonfluorescent phenotype as *ΔdnaKdnaJ* cells for both TGS (Figure 1d) and TG (Figure 1e), dnaJ259 cells expressing TGS and TG exhibited measurable fluorescence, albeit at a much lower level than was seen for wild-type cells (Figure 1f and g, respectively). Similar fluorescence results were observed when the TGS and TG constructs were expressed in AR7059 cells where the dnaJ gene was inactivated by transposon insertion (dnaJ::Tn10–42, see Table 1). Thus, it appears that the instability of Tat-targeted GFP in *ΔdnaKdnaJ* cells results predominantly from the deletion of dnaK, although involvement of DnaJ in this process cannot be ruled out.

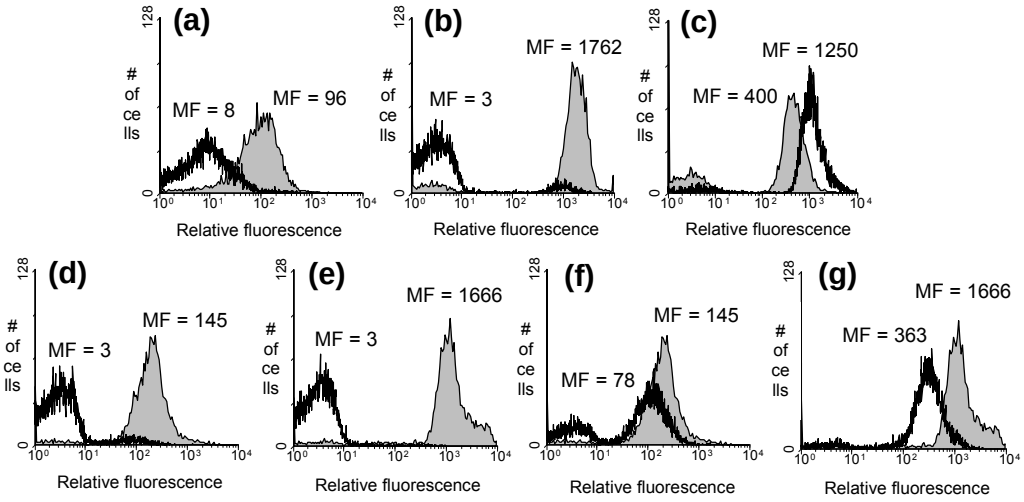


Figure 1. *Flow cytometric analysis of various E. coli strains expressing ssTorA-GFP-SsrA (TGS), ssTorA-GFP (TG) or GFP lacking a Tat signal (GFP).* Depicted are: strains MC4100 (gray) and CG3214 (as MC4100 *ΔdnaKdnaJ*; white) expressing (a) pMMB-TGS, (b) pTG or (c) pTrc-GFP; strains C600 (gray) and MF746 (as C600 dnaK756; white) expressing (d) pMMB-TGS or (e) pTG; and strains C600 (gray) and MF634 (as C600 dnaJ259; white) expressing (f) pMMB-TGS or (g)pTG. Mean fluorescence (MF) is the average of three replicate experiments where the standard error was <5%.

As mentioned above, cell fluorescence derived from TG expression is not an authentic indicator of Tat transport, since TG accumulates in both the cytoplasm and periplasm. Thus, to analyze the localization of TG in *ΔdnaKdnaJ* mutant cells more carefully, we performed subcellular fractionations. Specifically, TG was found to partition comparably between the cytoplasm and periplasm of wild-type MC4100 cells (Figure 2a), in agreement with numerous previous studies (Barrett et al., 2003; DeLisa et al., 2004; DeLisa et al., 2002; Santini et al., 2001; Thomas et al., 2001). In contrast, no detectable TG was present in fractions generated from *ΔdnaKdnaJ* cells, which was consistent with the lack of cell fluorescence seen above (Figure 1e). Since GFP expressed without a Tat signal peptide accumulated at a high level in the cytoplasmic fraction of *ΔdnaKdnaJ* cells (Figure 2a), we conclude

that DnaK is essential for the stability of GFP only when it carries an N-terminal Tat signal. To confirm whether this stability was specific to TG or a more general defect seen for other Tat substrates, we expressed a fusion between ssTorA and the *E. coli* maltose-binding protein (ssTorA-MBP) in wild-type and *ΔdnaKdnaJ* mutants. Previous reports have demonstrated successful routing of MBP through the Tat pathway (Blaudeck et al., 2003; Fisher et al., 2006; Tullman-Ercek et al., 2007). As was seen for TG, ssTorA-MBP partitioned normally in wild-type cells but was completely undetectable in the cytoplasmic and periplasmic fractions of *ΔdnaKdnaJ* cells (Figure 2b). As was observed for GFP, MBP expressed without an N-terminal signal peptide (ΔssMBP) was stable in the cytoplasm of *ΔdnaKdnaJ* cells (Figure 2b) confirming that the lack of detectable ssTorA-MBP in *ΔdnaKdnaJ* cells is somehow associated with Tat targeting. To determine whether the absence of soluble TG and ssTorA-MBP was due to accumulation in the membranes or insoluble fraction of cells, we probed whole cells for expression of each of these substrates. As seen in Figure 2c, wild-type cells but not *ΔdnaKdnaJ* mutants accumulated a significant level of TG and ssTorA-MBP, indicating that the lack of detectable Tat substrates in *ΔdnaKdnaJ* cells is not due to membrane stalling or partitioning into the insoluble fraction. Next, to ascertain whether the stability defect of ssTorA-MBP in *ΔdnaKdnaJ* cells required a functional Tat signal peptide, we expressed MBP carrying a variant TorA signal in which the quintessential twin arginine residues were replaced by twin lysine residues (ssTorA(KK)), a mutation that is known to block Tat transport (Cristobal et al., 1999). As expected, expression of ssTorA(KK)-GFP and ssTorA(KK)-MBP in wild-type cells resulted in accumulation of MBP in the cytoplasm only (Figure 2d). Interestingly, the same ssTorA(KK) constructs

were dramatically less stable in *ΔdnaKdnaJ* cells, as evidenced by the very faint cytoplasmic cross-reacting bands (Figure 2d), indicating that DnaK-mediated stability of Tat-targeted substrates does not require a fully-functional Tat signal peptide. Finally, to determine whether the general defect in Tat substrate stability caused by the absence of DnaK was due to impaired expression of the Tat machinery, we performed immunoblotting of whole cell fractions from wild-type and *ΔdnaKdnaJ* cells using TatA, TatB or TatC antiserum. As seen in Figure (e, TatA levels were unchanged in *ΔdnaKdnaJ* cells relative to wild-type cells. Likewise, the levels of TatB and TatC were comparable in wild-type and *ΔdnaKdnaJ* cells (data not shown).

Similar fractionation experiments were performed to determine the cellular distribution of TG and ssTorA-MBP in groEL and groES mutant strains. As was seen for *ΔdnaKdnaJ* cells, and consistent with the lack of cellular fluorescence (Table 1), groEL140 cells accumulated no detectable TG, whereas wild-type strain B178 accumulated significant levels of TG in the periplasm and, to a much lesser extent, in the cytoplasm (data not shown). Also in agreement with the cellular fluorescence data, we observed that the amount of TG that accrued in the groEL673 and groES30 strains was comparable to wild-type cells (data not shown). However, expression and localization of ssTorA-MBP in all the groEL and groES mutant strains was indistinguishable from that observed in wild-type cells (data not shown), suggesting that the effect of GroEL-GroES on TG and TGS was specific for GFP and not for Tat transport *per se*. Consequently, we focused our attention on the DnaK chaperone system.

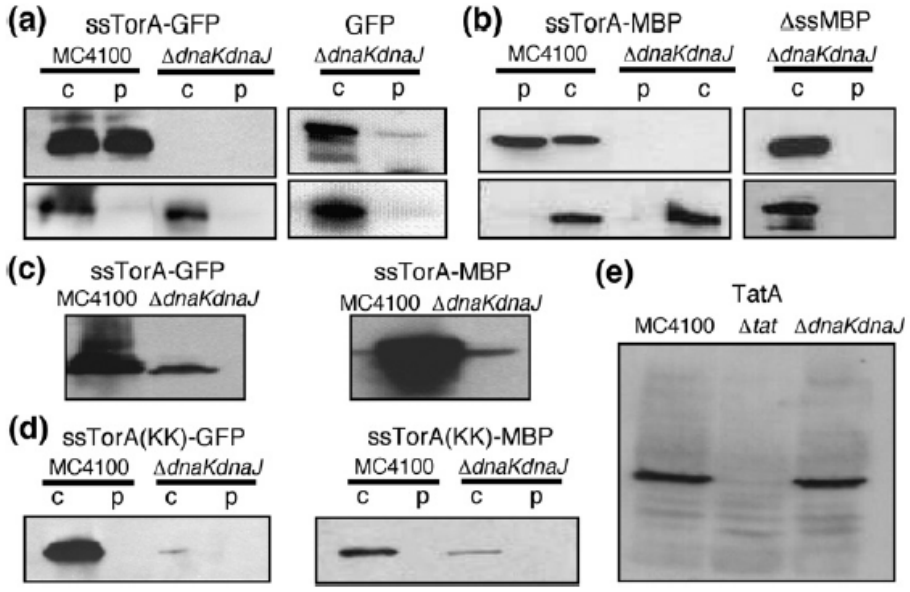


Figure 2. *Western blot analysis of ΔdnaKdnaJ mutant expressing heterologous Tat substrates.* (a) Subcellular fractionation of wt MC4100 and CG3214 (*ΔdnaKdnaJ*) cells expressing TG (left panels) or GFP lacking a signal peptide (right panels). Cytoplasmic (c) and periplasmic (p) samples were immunoreacted with monoclonal anti-GFP (upper panel) followed by monoclonal anti-GroEL (lower panel), the latter of which served as a fractionation marker.(b) Cytoplasmic (c) and periplasmic (p) fractions generated from wt MC4100 and *ΔdnaKdnaJ* cells expressing ssTorA-MBP (left panels) or ΔssMBP (right panels). Samples were immunoreacted with monoclonal anti-MBP (upper panel) followed by monoclonal anti-GroEL (lower panel). (c) Western blot analysis of whole cell lysates from cells expressing ssTorA-GFP (left panels) or ssTorA-MBP (right panels). Samples were generated by resuspending cells directly in SDS-PAGE loading buffer, heating at 100 °C, loading directly to SDS-PAGE gel and immunoreacting with anti-GFP or anti-MBP antibodies. (d) Western blot analysis of cytoplasmic (c) and periplasmic (p) fractions generated from wt MC4100 and *ΔdnaKdnaJ* cells expressing ssTorA(KK)-GFP (left panel) and ssTorA(KK)-MBP (right panel). (e) TatA expression levels assayed by Western blot analysis of membrane fractions generated from MC4100, DADE (Δ*tatABCD*Δ*E*) and CG3214 (*ΔdnaKdnaJ*). Samples were separated by electrophoresis, blotted onto nitrocellulose membranes and immunoreacted with anti-TatA serum. For all Western blots, an equivalent number of cells was loaded per each lane.

2.3. DnaK is essential for the stability of native, plasmid-encoded Tat substrates

Since our data indicate that DnaK (and, to a lesser extent, DnaJ and GrpE) has a prominent effect on the stability of the heterologous Tat substrates GFP and MBP, we next explored whether lack of DnaK or its co-chaperones affected the stability of native Tat substrates. For these experiments, two Tat substrates from *E. coli*, namely CueO, a multicopper oxidase, and SufI, a suppressor of the cell division defect caused by *ftsI*, were chosen because they: (1) are endogenous substrates of the *E. coli* Tat pathway, thereby ensuring that the data obtained are relevant to the export of physiological substrates; and (2) do not attach complex cofactors (Berks et al., 2005), thereby allowing us to assess the effect of DnaK without the complication of the cofactor insertion process. Overproduction of both CueO and SufI from plasmid pBAD33 resulted in the accumulation of each protein in the cytoplasm and periplasm of wild-type cells (Figure 3a and b, respectively). Further, the appearance of each substrate in the periplasmic space of wild-type cells but not in *ΔtatC* mutants (Figure 3a and b) confirmed the Tat-specific behavior of these proteins. Remarkably, neither CueO nor SufI was detectable in any fraction isolated from *ΔdnaKdnaJ* cells (Figure 3a and b), consistent with the instability observed for Tat-targeted GFP and MBP in this strain background. Also consistent with earlier results was the observation that both CueO and SufI were stably expressed and transported in cells carrying *dnaJ* and *grpE* mutations (Figure 3c).

To further demonstrate that DnaK-mediated stability of Tat substrates extends to any natural Tat substrate, even those from other bacteria that are compatible with the *E. coli* Tat system, we expressed *Thermus thermophilus*

alkaline phosphatase (Tap) from the pBAD24 vector in wild-type and several mutant strains. *T. thermophilus* Tap carries a canonical N-terminal Tat signal peptide that is capable of engaging the *E. coli* Tat machinery and routing the enzyme into the periplasm (Angelini et al., 2001). It was shown earlier and is confirmed here that Tap was not transported in a *ΔtatC* deletion strain and, as a result, was rapidly degraded so that no cellular Tap activity could be detected (Figure 3d). On the contrary, total cellular Tap activity in wild-type cells was high (Figure 3d) and this activity was split comparably between cytoplasmic and periplasmic fractions (data not shown). When Tap was expressed in *ΔdnaKdnaJ* or dnaK756 cells, no measurable Tap activity was observed. However, expression in dnaJ259, dnaJ::Tn10–42 or grpE280 cells resulted in cellular Tap activity that was nearly identical with that of wild-type cells (Figure 3d, results shown are for dnaJ259 only), consistent with our findings above that stability of over-expressed Tat substrates is regulated primarily by DnaK and, to a much lesser extent, by its co-chaperones DnaJ and GrpE. Collectively, these data indicate that DnaK is necessary for the stability of overproduced authentic Tat substrates in a manner that is independent of cofactor acquisition by the substrate.

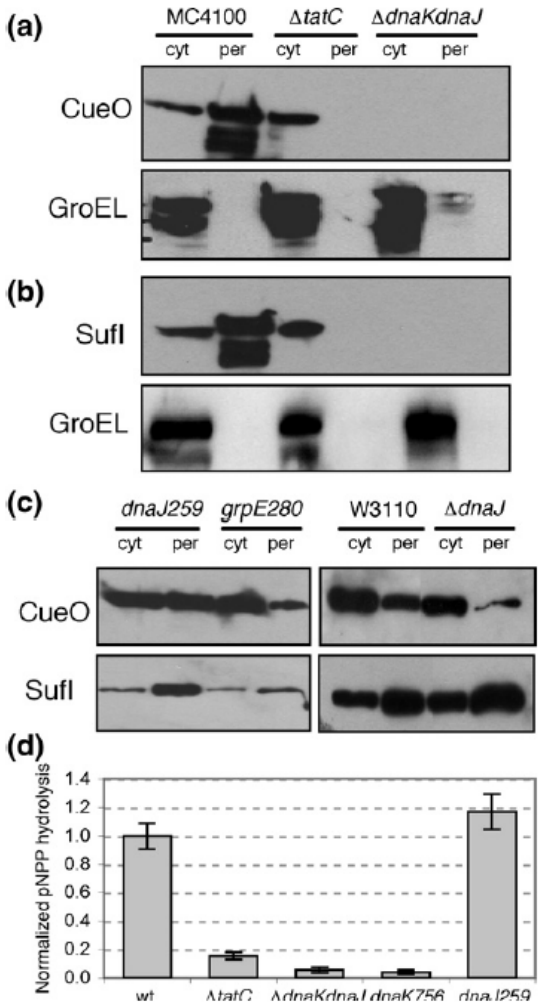


Figure 3. *Expression and localization of endogenous Tat substrates in ΔdnaKdnaJ mutant cells.* Western blot analysis of subcellular fractions generated from strains MC4100, B1LK0 (*ΔtatC*) and CG3214 (*ΔdnaKdnaJ*) expressing *E. coli* (a) CueO and (b) SufI. Cytoplasmic (cyt) and periplasmic (per) fractions were immunoreacted first with monoclonal anti-FLAG (upper panel) and then with anti-GroEL antibody (lower panel). An equivalent number of cells was loaded in each lane. (c) Western blot analysis of subcellular fractions generated from dnaJ259 and grpE280 mutants (left panels) and W3110 and dnaJ::Tn10–42 (ΔdnaJ) cells (right panels) each expressing CueO (upper panels) or SufI (lower panels). (d) Expression of *T. thermophilus* Tap in MC4100 (wt), B1LK0 (*ΔtatC*), CG3214 (*ΔdnaKdnaJ*), MF746 (dnaK756) and MF634 (dnaJ259) cells. Activity was assayed by monitoring pNPP hydrolysis at 405 nm. Data reported are the average of three replicate experiments and each sample was assayed a minimum of three times.

2.4. DnaK maintains efficient transport of chromosomally encoded Tat substrates but is not essential for stability

To test the effect of DnaK on physiological Tat transport, we assayed for subcellular localization and activity of the endogenous Tat substrate trimethylamine N-oxide reductase (TorA) expressed from the chromosome of *ΔdnaKdnaJ* cells. TorA is a monomeric protein that attaches a molybdopterin guanine dinucleotide cofactor in the cytoplasm and is subsequently translocated into the periplasm, where it remains as a water-soluble enzyme that facilitates anaerobic respiration with trimethylamine N-oxide (TMAO) as terminal electron acceptor. Following subcellular fractionations of cells grown anaerobically with TMAO, we found that the total TMAO activity was nearly identical in wild-type and *ΔdnaKdnaJ* cells and, further, that the partitioning of TMAO activity in wild-type and *ΔdnaKdnaJ* cells was nearly identical (Figure 4a). We tested *ΔdnaKdnaJ* cells for their ability to translocate a second chromosomally encoded Tat substrate; namely, formate dehydrogenase N subunit G (FdnG). FdnG carries a Tat-dependent signal peptide, attaches complex cofactors (e.g. 4Fe-4S) and forms a ternary complex along with the FdnH and FdnI subunits that together are localized to the periplasmic side of the inner membrane. Similar to chromosomally expressed TorA, Western blot analysis revealed that localization of the FdnG protein was unaffected in*ΔdnaKdnaJ* mutant cells relative to wild-type cells, as both accumulated FdnG predominantly in the membrane fraction (data not shown). Total formate dehydrogenase N activity was reproducibly lower in *ΔdnaKdnaJ* cells relative to wild-type cells (Figure 4b), suggesting that either translocation efficiency of the enzyme complex was slightly impaired or enzyme assembly

(i.e., cofactor attachment to FdnG or FdnGHI oligomerization) was partially disrupted in the *ΔdnaKdnaJ* background.

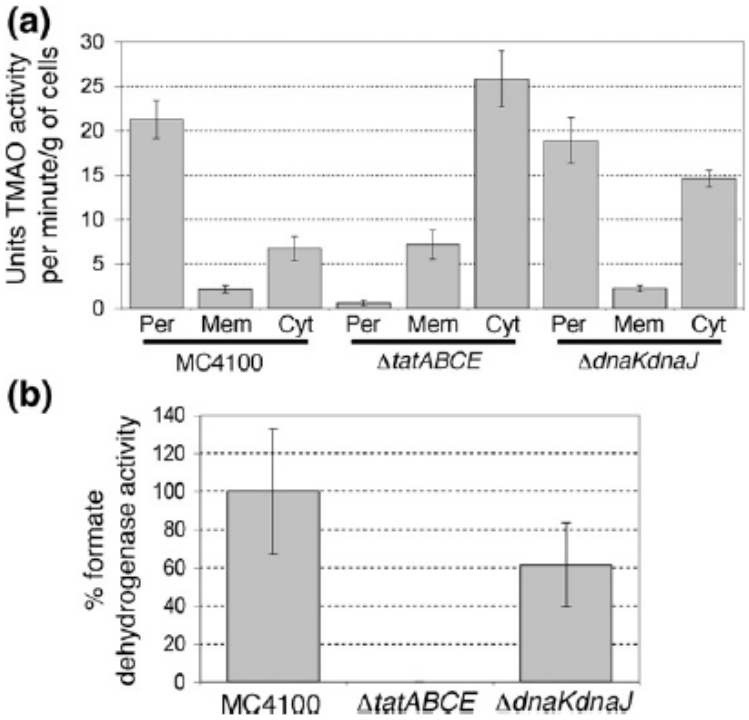


Figure 4. *Subcellular localization of chromosomally encoded E. coli Tat substrates in strains lacking functional DnaK-DnaJ*. (a) Cytoplasmic (cyt), membrane (mem) and periplasmic (per) fractions of MC4100, DADE (*ΔtatABCDΔE*) and CG3214 (*ΔdnaKdnaJ*) cells were assayed for TMAO reductase activity and normalized per gram of cells. (b) Formate dehydrogenase (FdnG) activity present in lysate from MC4100, DADE (*ΔtatABCDΔE*) and CG3214 (*ΔdnaKdnaJ*). FdnG activity was normalized to the activity found in MC4100 cells and is reported as a percentage of this wt activity. All data reported are the average of three replicate experiments.

2.5. DnaK crosslinks to the TorA502 protein in vivo

Since the effect of DnaK on the stability of overexpressed Tat substrates was not due to altered levels of TatABC expression (see Figure 2e), we reasoned that DnaK might interact with these substrates directly. The rationale for this was an earlier report that DnaK was one of only two proteins (the other was the REMP DmsD) that bound to a Tat-specific signal peptide *in vitro* (Oresnik et al., 2001). Thus, to determine whether DnaK associates with Tat substrates *in vivo*, we performed crosslinking experiments using cell lysate generated from wild-type cells expressing a truncated version of TorA that lacks nine molybdo-cofactor ligands essential for cofactor attachment (Chanal et al., 2003). TorA502 was chosen because: (1) lack of functional DnaK resulted in decreased wild-type TorA transport efficiency (see Figure 4(a)); (2) proteins bearing the TorA signal peptide often saturate their own translocation (Chanal et al., 2003), a phenomenon that we hypothesize requires DnaK to overcome; and (3) the targeting of TorA502 to the Tat pathway is conveniently decoupled from the cofactor insertion process. It should be noted that TorA502 is not translocated across the inner membrane; however, its targeting to the Tat machinery has been confirmed by its ability to inhibit the translocation of full-length TorA (Chanal et al., 2003). That is, overproduction of TorA502 from pBAD24 results in saturation of the Tat machinery such that full-length TorA expressed from the chromosome is not detected in the periplasm (Figure 5(a)). Overproduction of TorA502 lacking its N-terminal signal peptide (ΔssTorA502) relieves the saturation and results in periplasmic accumulation of full-length TorA (Figure 5b). Interestingly, full-length TorA was never found in the periplasm of *ΔdnaKdnaJ* cells in these experiments, regardless of whether TorA502 carried a signal peptide (compare Figure 5a

and b), providing further evidence that chromosomally encoded Tat substrates are poorly translocated in the absence of a functional DnaK system. Note that only the full-length TorA band is shown in Figure 5a and b and, while the lower molecular mass TorA502 was also readily detected by the anti-TorA serum, the relatively longer development time needed to visualize the full-length TorA band caused the TorA502 band to become greatly over-exposed (data not shown).

Importantly, when cell lysate from wild-type cells was treated with formaldehyde to induce crosslinking, a faint but reproducible anti-TorA crossreacting band appeared at a molecular mass of ~120 kDa (Figure 5c, TD*), consistent with the expected size of a TorA502-DnaK complex. Probing of the same samples with monoclonal anti-DnaK revealed an identical ~120 kDa band (Figure 5d), confirming that this high-molecular mass complex likely corresponds to a TorA502-DnaK complex. Unexpectedly, when we further probed these same samples with monoclonal anti-GroEL, a single high molecular mass product of ~110 kDa appeared (Figure 5e, TG*) that corresponded exactly to one of the crosslinked products that emerged using anti-TorA serum. Thus, it appears that both DnaK and GroEL can associate with at least one Tat-targeted substrate *in vivo*. Finally, to determine if these observed crosslinks were dependent upon the TorA502 signal peptide, we performed an identical experiment using cell lysate generated from wild- type cells expressing ΔssTorA502. Following treatment with formaldehyde and probing with TorA, DnaK and GroEL antibodies, we observed the same pattern of crosslinking (data not shown), indicating that the N-terminal signal peptide is not essential for binding of DnaK or GroEL to TorA502. This is not entirely surprising, as there are two predicted DnaK binding sites within the

TorA signal peptide (one of which overlaps the S/T-R-R-x-F-L-K consensus motif) but 14 putative DnaK-binding sites throughout the remainder of the TorA502 sequence as determined using an algorithm developed by Bukau et al. (Rudiger et al., 1997). Nonetheless, while the TorA signal peptide was not essential for DnaK (or GroEL) binding, we cannot rule out the possibility that chaperone binding sites that occur early in TorA and other Tat substrates might serve as specific signals for the timing and specificity of DnaK binding. In support of this notion, while *E. coli* Tat substrates have, on average, a slightly lower frequency of total DnaK-binding sites relative to their Sec counterparts or to cytoplasmic proteins (Figure 5f), the average position of the first DnaK-binding site in Tat substrates occurs markedly later in their primary sequence relative to Sec pathway substrates but much earlier compared to cytoplasmic proteins (Figure 5g).

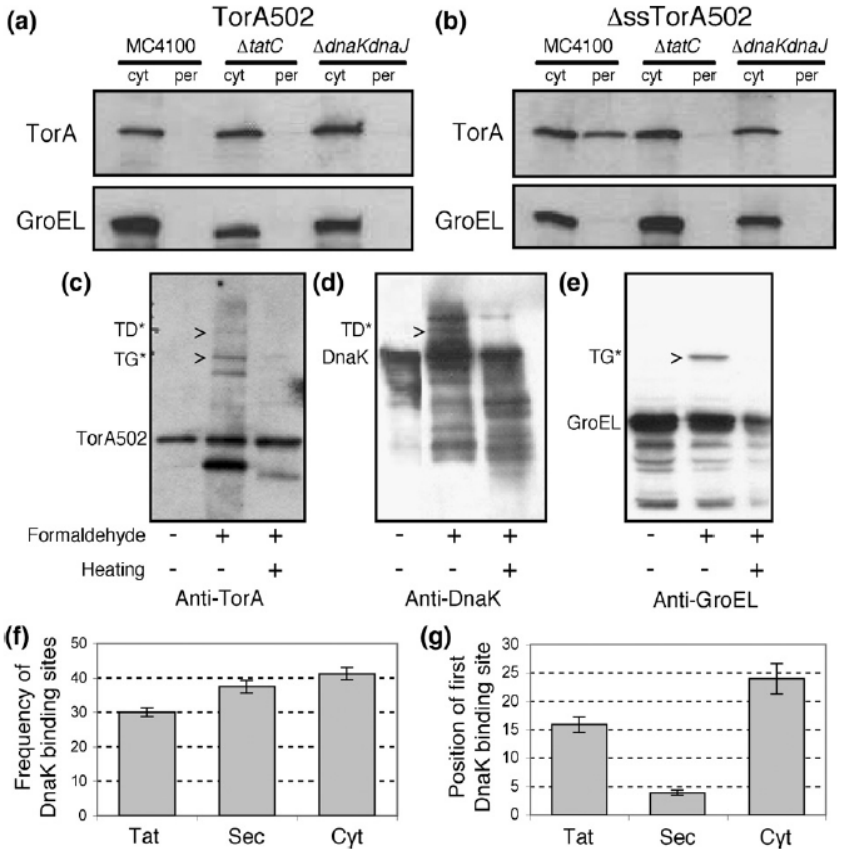


Figure 5. *In vivo crosslinking analysis of truncated TorA with DnaK*. Western blot analysis of cytoplasmic (cyt) and periplasmic (per) fractions generated from wt MC4100 cells expressing (a) TorA502 or (b) ΔssTorA502. The upper panels are blots that were immunoreacted with anti-TorA serum while the lower panels were immunoreacted with anti-GroEL antibody. The bands seen in the upper panels depict a single 90 kDa band that corresponds to full-length TorA expressed endogenously. Note that a second group of more highly expressed, lower molecular mass bands was detected in the cytoplasm of all fractions (not shown), among them the biggest was ~60-kDa and corresponded to TorA502 expressed from the plasmid. Lysate from wt MC4100 cells expressing TorA502 was treated as indicated and tested for crosslinking as follows: (c) blots immunoreacted with anti-TorA serum. Note that for the exposure time shown here, the endogenously expressed TorA that was seen in (a) is no longer visible and that the band seen in the first lane corresponds to TorA502. The same samples probed with (d) anti-DnaK and (e) anti-GroEL monoclonal antibodies, respectively. Bioinformatic prediction of (f) the total number of DnaK-binding sites and (g) average position of the first putative DnaK binding site in Tat, Sec and cytoplasmic proteins. In all, 29 putative Tat substrates, 30 randomly selected Sec proteins, and 30 randomly selected cytoplasmic proteins were analyzed. The total number of DnaK binding sites for each protein was normalized by the number of total amino acids in that protein's primary sequence.

2.6. Increased cellular levels of DnaK-DnaJ promotes high-level periplasmic accumulation of Tat substrates.

On the basis of the fact that (1) a collection of recombinantly expressed Tat substrates were unstable in cells lacking DnaK and (2) DnaK was found associated with a Tat substrate *in vivo*, we hypothesized that co-expression of DnaK (and its co-chaperone DnaJ) would markedly improve the half-life of Tat substrates in the cytoplasm leading to a concomitant increase in the steady-state accumulation of these substrates in the periplasmic space. To test this notion, we first expressed TG in an *E. coli* strain where the expression of the chromosomal dnaKdnaJ operon is under the control of the IPTG inducible promoter PA1/lacO-1 (referred to here as P_{IPTG}*dnaKJ*) (Deuerling et al., 1999). In these cells, the expression of *dnaKdnaJ* is regulated by IPTG such that the DnaK and DnaJ levels are adjustable from <10% of wild-type levels in the absence of IPTG to about 200% in IPTG (Tomoyasu et al., 1998). Treatment of P_{IPTG}*dnaKJ* cells with IPTG before TG expression resulted in a much higher level of TG in the periplasm relative to wild-type cells, as revealed by Western blot analysis and activity assays (Figure 6a). This increase in periplasmic TG correlated with a large decrease in the level and activity of TG in the cytoplasm. A Western blot using monoclonal anti-DnaK confirmed that the level of DnaK increased with the addition of IPTG to a level that was approximately three to four times above that observed in wild-type cells (Figure 6a). Thus, it appears that by up-modulating the concentration of DnaK-DnaJ in the cytoplasm, we were able to improve the transport efficiency of TG in P_{IPTG}*dnaKJ* cells.

Next, we explored whether higher levels of cellular DnaK-DnaJ would further improve the cytoplasmic stability and translocation of TG. For this

experiment, we transformed wild-type cells carrying pTG with a second plasmid, pOFXbad-KJ2, which expresses *dnaKdnaJ* from the araBAD promoter (Castanie et al., 1997). Upon full induction of TG and *dnaKdnaJ* with 0.2% arabinose, we observed a dramatic increase in the total expression level of TG relative to that in wild-type cells carrying pTG and empty pOFXbad2, with a significant portion of this protein localizing in the periplasm (Figure 6b). Western blot analysis using monoclonal anti-DnaK confirmed the massive increase in cellular levels of DnaK for cells expressing pOFXbad-KJ2. In agreement with Western blot results, activity measurements indicated a ~3-fold increase in active periplasmic TG (Figure 6b). In contrast to the results obtained using P_{IPTG}*dnaKJ* cells, when DnaKDnaJ were greatly over-expressed, the relative abundance of TG increased dramatically in both the cytoplasm and periplasm; however, the relative partitioning between these two compartments remained nearly unchanged. For comparison, we separately co-transformed pTG-carrying cells with pOFXbad-SL2 that expresses *groEL* and *groES* from behind the araBAD promoter and found that coexpression of GroEL-GroES resulted in higher total expression of TG in both the cytoplasm and the periplasm relative to control cells, as determined by Western blot analysis and activity measurements (data not shown).

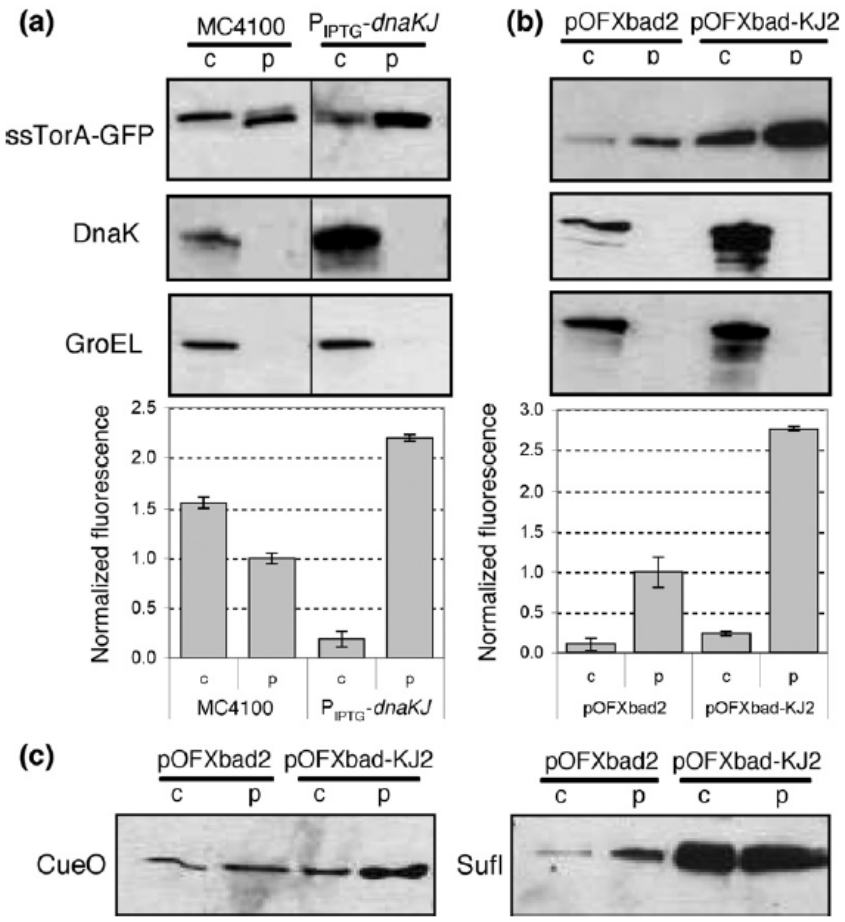


Figure 6. *Effect of increased levels of DnaK-DnaJ on Tat transport.*(a) Western blot analysis and fluorescence activity assay of cytoplasmic (c) and periplasmic (p) fractions generated from MC4100 and P_{IPTG}*dnaKJ* cells expressing ssTorA-GFP (TG). Expression of TG was induced with 0.2% (w/v) arabinose while *dnaKJ* was induced with 1.0 mM IPTG. The upper, middle and lower panels depict membranes that were immunoreacted with anti-GFP, anti-DnaK and anti-GroEL, respectively.(b) Western blot analysis and fluorescence activity of cytoplasmic (c) and periplasmic (p) fractions generated from MC4100 co-expressing TG and either pOFXbad2 (empty control) or pOFXbad-KJ2 as indicated. Co-expression of TG and *dnaKJ* was induced with 0.2% arabinose. (c) Western blot analysis of MC4100 cells co-expressing CueO or SufI and either pOFXbad2 or pOFXbad-KJ2 as indicated. Co-expression of CueO (or SufI) and *dnaKJ* was induced with 0.2% arabinose. Blots were probed with anti-FLAG. Note that an equivalent number of cells was loaded in each lane for all blots shown and all GFP activity data reported are the average of three replicate experiments and each sample was assayed a minimum of three times

Finally, we explored whether DnaK-DnaJ coexpression could similarly improve the expression and/or translocation of overproduced *E. coli* Tat substrates such as CueO or SufI. Recall that each of these substrates was dependent on DnaK for stability (see Figure 3). When we co-expressed DnaK-DnaJ from pOFXbad2 along with CueO or SufI, we observed a moderate improvement in the steady-state periplasmic accumulation of CueO and a more dramatic improvement in the case of SufI (Figure 6c).

In addition, the total expression level (cytoplasm + periplasm) of both CueO and SufI was significantly higher when DnaK-DnaJ were overproduced compared to cells expressing normal levels of DnaK-DnaJ. This latter observation suggests that DnaK-DnaJ-mediated improvement in Tat substrate expression and translocation is due, in part, to enhanced stability conferred by the chaperone pair. Interestingly, and in agreement with its less consistent role in the Tat transport process, coexpression of GroEL-GroES had no measurable affect on CueO or SufI expression and translocation (data not shown).

3. Discussion

We undertook the experiments presented here to determine whether previously undetected auxiliary factors, in addition to the TatABC machinery, participate in Tat transport. Numerous lines of evidence suggested to us that general molecular chaperones (e.g. DnaK, GroEL) and proteases (e.g., ClpXP, FtsH) play an important role in the early stages of Tat targeting and translocation in bacteria (DeLisa et al., 2004; Oresnik et al., 2001; Rodrigue et al., 1996; Xu et al., 2005) and plants (Dionisi et al., 1998; Leheny et al., 1998;

Madueno et al., 1993; Molik et al., 2001) (for a review see Fisher and DeLisa (Fisher and DeLisa, 2004). Accordingly, we tested a panel of *E. coli* mutant strains lacking well-characterized chaperones or proteases and found that the molecular chaperone DnaK exerted the most significant effect on the expression and transport of synthetic Tat substrates such as ssTorA-GFP-SsrA, ssTorA-GFP and ssTorA-MBP. Interestingly, the loss of DnaK invariably rendered these substrates completely unstable, whereas the loss of functional DnaJ or GrpE resulted in only a minor stability defect, suggesting that the participation of DnaK in Tat substrate stability is not entirely dependent on its co-chaperones. In the case of DnaJ, it is possible that an alternate co-chaperone such as DjlA or CbpA may substitute for DnaJ (Gur et al., 2004) in mediating Tat substrate stability. In the case of GrpE, since we were only able to test a conditionally lethal mutant (*grpE280*) and not a deletion mutant (*grpE* is an essential gene), it is plausible that the defect in GrpE280 does not hinder its participation in stabilizing Tat substrates. Thus, while we did not observe a major stability defect for *dnaJ* and *grpE* mutants directly, we are unable to rule out their participation in this phenotype. In addition to affecting the stability of these engineered reporters constructs, the absence of DnaK resulted in a lack of stabilization for the endogenous Tat substrates CueO and SufI when each was overproduced from a pBAD expression vector. Interestingly, chromosomally-encoded Tat substrates FdnG and TorA were expressed stably and translocated in strains lacking DnaK, although transport efficiency was decreased relative to that of wild-type cells. One explanation for why the loss of DnaK was less significant on physiological expression of FdnG and TorA is that both of these enzymes have cognate REMPs (Palmer et al., 2005) that may render these substrates insensitive to the loss of DnaK under normal

levels of expression. Nonetheless, taken together, our data suggest that DnaK has a direct and significant role in stabilizing Tat substrates before engaging the TatABC translocon and in maintaining efficient Tat export. Consistent with this role, previous studies have found that DnaK binds specifically to Tat signal peptides (Oresnik et al., 2001) and to Tat system-specific chaperones such as DmsD. It is noteworthy that while GroEL-GroES emerged in our initial genetic screen, follow-up studies revealed highly inconsistent involvement of GroEL-GroES in the transport of different Tat substrates, which is likely an outcome of this chaperone's broad substrate specificity and ability to assist the folding of a relatively large number of cytoplasmic proteins (Erbse et al., 2003; Houry et al., 1999; Kerner et al., 2005) including a subset of Tat substrates (Kerner et al., 2005; Madueno et al., 1993; Molik et al., 2001; Rodrigue et al., 1996).

In light of its significant role in maintaining Tat substrate stability, it is intriguing to speculate how DnaK might accomplish this role in the bacterial cytoplasm. First and foremost, even though DnaK has been linked only to cofactor-containing Tat substrates, (Dionisi et al., 1998; Madueno et al., 1993; Oresnik et al., 2001; Rodrigue et al., 1996) the observed instability of Tat substrates lacking cofactors (e.g. CueO, SufI, GFP and MBP) in the absence of DnaK suggests that the role of these chaperones in Tat substrate stability is independent of cofactor biogenesis. Rather, we favor a model whereby DnaK binding to Tat substrates, as was observed for TorA502 (see Figure 5), serves to shield Tat substrates effectively from degradation by proteases and premature interaction with the TatABC machinery. Indeed, it has been observed repeatedly that rapid proteolytic degradation eliminates Tat substrates that are rendered incompetent for export due to (1) non-functional

signal peptides such as RR→KK variants (our unpublished results), (2) improper or incomplete folding (DeLisa et al., 2003; Fisher et al., 2006) or (3) lack of functional TatABC machinery (Angelini et al., 2001; Chanal et al., 1998; Santini et al., 2001). Such a mechanism would be particularly important under conditions where Tat substrates become backlogged in the cytoplasm, a phenomenon that occurs commonly when Tat substrates are overproduced from expression vectors (Chanal et al., 2003; DeLisa et al., 2004; Yahr and Wickner, 2001). Under circumstances where a fully export-competent Tat substrate is delayed for transport due to translocon saturation, it is possible that DnaK may prevent untimely cytoplasmic degradation. A further possibility is that overproduced Tat substrates that aggregate before transport may be rescued by DnaK-mediated shuttling to anti-aggregation factors such as ClpB (Mogk et al., 2003). Interestingly, expression and translocation of ssTorAGFP-SsrA and ssTorA-GFP was negatively affected by ClpB deletion (Table 1). It is possible, although not explicitly supported by any of the data presented here, that DnaK participates in substrate targeting to the TatABC translocon.

Another important question is how DnaK mediated stability of Tat substrates fits within the cellular folding-maturation cascade that begins for most proteins during translation on the ribosome and ends with a fully folded protein. Luirink and coworkers used site-specific photo-crosslinking to demonstrate that TF contacts the signal peptide of nascent Tat proteins early in biogenesis; however, and in agreement with our data (Table 1), depletion or over-expression of TF had little effect on the kinetics and efficiency of the Tat export process (Jong et al., 2004). Following this early interaction with TF, the next likely chaperones to engage Tat substrates are REMPs (Jack et al., 2004). Since the first putative DnaK-binding site in most Tat substrates occurs

in and around the S/T-R-R-x-F-L-K consensus motif (see Figure 5), we postulate that DnaK and REMPs may compete for overlapping binding sites in Tat signal peptides. Alternatively, DnaK and REMPs may communicate with one another to coordinate their interaction with the substrate. In support of this latter notion, it has been observed recently that the REMPs DmsD and NapD associate with the molecular chaperones DnaK and GroEL (Butland et al., 2005), although the physiological significance of these interactions remains to be demonstrated. Surprisingly, although over-expression of the REMP TorD has been shown to improve TG expression and transport, (Li et al., 2006) none of the REMPs tested in our experiments (DmsD and TorD) had any effect on ssTorA-GFP-SsrA or ssTorA-GFP, although it is possible that their participation in the translocation of these proteins may have escaped detection by our flow cytometric screening approach. Collectively, an emerging picture of Tat substrate targeting involves the highly coordinated participation of Tat specific as well as general molecular chaperones that function to stabilize the substrate against misfolding and proteolysis as well as help to escort Tat substrates to the TatABC translocon while preventing early engagement of the Tat machinery.

Finally, the observation that co-expression of the DnaK-DnaJ chaperones was able to improve the expression and translocation efficiency of several Tat-targeted substrates further strengthened the connection between DnaK and the Tat export process. We speculate that this improvement is a direct result of: (1) the protection of Tat substrates from cytoplasmic degradation or premature translocon engagement; and (2) an increase in overall folding yield achieved by Tat substrates such that a greater fraction of overproduced Tat substrates are in an export competent (i.e., correctly folded)

conformation. In the case of GFP, improved Tat transport appears to arise predominantly by the latter mechanism as the soluble yield of GFP lacking a Tat signal was increased nearly as much as ssTorA-GFP when DnaK-DnaJ were co-expressed from a plasmid (data not shown). From a biotechnology perspective, these results are significant, as they provide a general strategy for increasing the utility of the Tat system as a platform for applications that hinge on protein expression and secretion.

4. Materials and Methods

4.1. Bacterial strains, plasmids and growth conditions

All strains and plasmids used in this study are listed in Tables 2 and 3, respectively. *E. coli* strain MC4100 (F- *araD139* ∆(*argF-lacU169*) *rpsL150 relA1 flbB5301 deoC1 ptsF25 rbsR*) or its derivatives were used for all experiments unless otherwise noted. To confirm the stability of CueO and SufI without their natural Tat signal peptides, the *E. coli cueO* and *sufI* genes were cloned with a C-terminal FLAG affinity tag into pBAD33 (Guzman et al., 1995). The ΔssCueO-FLAG construct was generated *via* PCR using pCueO-FLAG as a template (DeLisa et al., 2004) and primers that incorporated *Xba*I and *Hin*dIII restriction sites at the 5' and 3' ends, respectively. The PCR product was gel-purified, digested with *Xba*I and *Hin*dIII, and cloned into the corresponding sites of pBAD33, resulting in pΔssCueO-FLAG. The ΔssSufI-FLAG sequence was similarly constructed using pSufI-FLAG as template (DeLisa et al., 2004) and primers that incorporated *Sac*I and *Xba*I restriction sites at the 5' and 3' ends, respectively. The digested and purified PCR product was ligated into the *Sac*I and *Xba*I sites of pBAD33, resulting in pΔssSufI-FLAG. Plasmids for the

expression of MBP were constructed by PCR amplification of mature *E. coli malE* lacking its native Sec signal peptide. During PCR, *Xba*I and *Hin*dIII restrictions sites were introduced at the 5' and 3' ends, respectively. The resulting *Xba*I-*Hin*dIII-digested PCR product was ligated into the corresponding sites of pTrc99A (Amersham Pharmacia) yielding pΔssMBP. For the construction of pTorA-MBP, plasmid pTorA-cassette (pTrc99A containing the DNA sequence for the ssTorA signal peptide between *Nco*I and *Xba*I) was digested with *Xba*I and *Hin*dIII and the PCR product encoding mature MBP was ligated into this position. Generation of pTorA(KK)-MBP was accomplished using site-directed mutagenesis (Stratagene QuickChange® site-directed mutagenesis kit) to change the consensus arginine to twin lysine residues.

Strains were routinely grown aerobically at 37 °C in Luria-Bertani (LB) medium, and antibiotic supplements were at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 25 μg/ml; tetracycline 10 μg/mL. Protein synthesis was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG; 0.1 mM) and/or arabinose (0.2% w/v) when the cells reached an absorbance at 600 nm of ~0.5.

4.2. Flow cytometric analysis

Overnight cultures harboring GFP-based plasmids were subcultured into fresh LB medium with chloramphenicol and induced with 0.1 mM IPTG (pMMB-TGS and pTrc-TG) or 0.2% arabinose (pTG) in mid-exponential phase growth. After 4-6 h, cells were washed once with phosphate-buffered saline, and 5 μl of washed cells were diluted into 1 ml of phosphate-buffered saline prior to analysis using a Becton Dickinson Biosciences FACSCalibur.

4.3. Subcellular fractionation

The fraction of periplasmic proteins was obtained using the ice-cold osmotic shock procedure (DeLisa et al., 2003; Sargent et al., 1998). The quality of all fractionations was determined by immunodetection of the cytoplasmic GroEL protein (DeLisa et al., 2003). The whole cell lysate fraction was prepared by first washing the pellet in PBS followed by centrifugation and resuspension of the pellet directly in 250 μL of SDS-PAGE loading buffer and finally heating for 10-15min at 100°C. For TorA and FdnG experiments, membrane fractions were prepared by lysing sphaeroplasts prepared by the lysozyme/EDTA method (Stanley et al., 2001) using a French press, followed by ultracentrifugation at 120,000x*g* for 30 minutes.

4.4. Western blot analysis

Proteins were separated by SDS-PAGE and Western blotting was as described (Chen et al., 2001). The following primary antibodies were used with the corresponding dilution factors in parenthesis: mouse anti-GFP (1:2,000; Sigma), mouse anti-MBP (1:4,000; Sigma); mouse anti-DnaK (1:1,000; Stressgen), mouse anti-GroEL (1:20,000; Sigma); mouse anti-FLAG (1:1,000; Stratagene); rabbit anti-TorA (1:5,000; Palmer laboratory stock) and rabbit anti-FDH-N (1:2,000; kindly provided by Professor D.Boxer, University of Dundee, UK). Immunoblotting against TatA and TatB components in membrane fractions was carried out as described in De Leeuw *et al.* (De Leeuw et al., 2001) and against TatC as described by Alami *et al.* (Alami et al., 2002). The secondary antibodies were goat anti-mouse and goat anti-rabbit horseradish peroxidase conjugates diluted 1:10,000. To verify the quality of subcellular fractions, membranes were first probed with primary

antibodies and, following development, were stripped in Tris-buffered saline/2% SDS/0.7 M β-mercaptoethanol. Stripped membranes were reblocked and probed with anti-GroEL antibody.

4.5. Protein activity assays

Data reported for all activity assays was the average of at least 3 replicates. Alkaline phosphatase activity was quantified by monitoring the rate of hydrolysis of *p*-nitrophenyl phosphate (pNPP, Sigma) as described previously (Derman et al., 1993). GFP activity in subcellular fractions was quantified by diluting samples into PBS in a 96-well plate and measuring the emission at 509 nm using a fluorescence microplate reader (BioTek Instruments). TMAO reductase activity assays were performed by growing cells anaerobically on LB supplemented with 0.4% (w/v) glucose and 0.4% (w/v) TMAO followed by fractionation as detailed above. TMAO activity was measured as described (Horvath, 2007; 1988). FdnG activity was measured using the method of Enoch and Lester (Enoch and Lester, 1975). Briefly, 10µl of 22mg/ml phenazine methosulphate (PMS) was added to a 1.8 ml glass cuvette with a sealing stopper, bored to accommodate a Hamilton syringe. 4 or 5 2-mm glass balls were then added to facilitate mixing of the contents. To this, 50 mM anaerobic Tris-Hcl (pH8.0) was added to fill the cuvette, which was then sealed. A sufficient amount of 2,6-dichlorophenol-indophenol (DcPiP) solution (10 mg/ml) was then added using a Hamilton syringe to give an absorbance of 1.0 at 600 nm. Material from 1 mg of cells grown in LB medium supplemented with 0.4% (w/v) KNO_3 in place of Cohen and Rickenberg medium was then added. Upon observing a steady absorbance at 600 nm, 10 µl of 20% (w/v) sodium formate was added to start the reaction.

Enzyme activity was measured by following the reduction in absorbance at 600 nm over time and calculated using the molar extinction coefficient of DcPip at 600 nm of 21,000.

4.6. Crosslinking analysis

Cytoplasmic fractions prepared from wild-type cells carrying pTorA502 (in phosphate buffer) were treated with 0.6% (v/v) formaldehyde. The reaction was incubated at room temperature for 30 min and was stopped by the addition of SDS-PAGE loading buffer. To dissociate cross-linked complexes, the reaction was heated at 100°C for 15 min, whereas the control was heated at 100°C for 5 min with SDS charge blue buffer in the presence of β-mercaptoethanol before loading to the SDS/polyacrylamide gel.

4.7. Bioinformatic prediction of DnaK binding sites

A bioinformatics algorithm for predicting DnaK binding sites in primary sequences of proteins was kindly provided by Dr. Bernd Bukau and has been previously described in detail (Rudiger et al., 1997). All 29 putative *E. coli* Tat substrates (Tullman-Ercek et al., 2006) were analyzed using this algorithm to determine the total number of predicted DnaK binding sites as well as the relative location of the first binding site along the primary sequence of each substrate. For comparison, 30 Sec-dependent proteins and 30 cytoplasmic proteins were chosen at random and analyzed in an identical manner. Finally, the number of total DnaK binding sites for each protein analyzed was normalized by the total length of that particular protein.

5. Acknowledgements

We thank Bernd Bukau for kindly providing strains, plasmids and the DnaK binding site algorithm used in these studies. We thank Francois Baneyx, Joseph Peters, Frank Sargent and Raymond Turner for their generous gifts of strains and antibodies. R.P.-R. gratefully acknowledges Cornell University for support through a Sloan Fellowship. This material is based upon work supported by the National Science Foundation under grant no. 0449080 (to M.P.D.), the National Institutes of Health under grant no. AI063709 (to M.P.D), a NYSTAR James D.Watson Award (to M.P.D.), LSHB-CT-2004- 005257 (to L.-F.W.) and the Medical Research Council for supporting T.P. and M.G.H. via the award of a Senior Non Clinical Research Fellowship (to T.P.).

CHAPTER 3

ENVELOPE STRESS IS A TRIGGER OF CRISPR RNA-MEDIATED DNA SILENCING IN PROKARYOTES[3]

1. Introduction

The use of small RNAs to regulate gene expression is ubiquitous in all living organisms including bacteria (Waters and Storz, 2009). In one remarkable instance, bacteria and archaea acquire resistance to conjugative plasmids, transposable elements and phages by employing an RNA-mediated defense mechanism against these foreign invaders. In this process, short fragments (~24 to 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 (Barrangou et al., 2007; Bolotin et al., 2004; Brouns et al., 2008; Lillestol et al., 2006; Makarova et al., 2006; Sorek et al., 2008). CRISPR loci have been identified in ~40% of eubacterial genomes and nearly all archaeal genomes sequenced to date (Sorek et al., 2008). They are often adjacent to genes encoding CRISPR-associated (Cas) proteins (Figure 7a). It is now firmly established that integration of phage-derived spacers into the genomic CRISPR loci confers resistance to the phages and the phage-resistant phenotype is dependent upon several of the *cas* gene products (Barrangou et al., 2007; Brouns et al., 2008). These findings are consistent with the hypothesis that CRISPR-Cas system encodes a nucleic acid-based 'immunity' mechanism in which the CRISPR spacer sequences

[3] Submitted to the Cell Journal.

govern target specificity and the Cas machinery provides the resistance (Makarova et al., 2006).

In the model proposed by Makarova and coworkers, RNA derived from spacers guide the Cas proteins to foreign genetic elements by base-pairing with either template or coding strands of the target, resulting in subsequent silencing of the target by additional Cas proteins. Although incomplete, the molecular details of this system are starting to come into focus. For instance, it appears that CRISPRs are initially transcribed into multiples of the CRISPR repeat unit including the intervening spacer sized about 500 nucleotides (Tang et al., 2002). These transcripts are subsequently processed to precursor RNAs (pre-crRNA) that contain at least two or three repeat-spacer units (~120 and ~180 nucleotides) (Brouns et al., 2008; Hale et al., 2008). These pre-crRNAs are processed into non-coding small CRISPR RNA (crRNA) products that are ~50-70 nucleotides long (Brouns et al., 2008; Hale et al., 2008; Lillestol et al., 2006; Tang et al., 2002). In the archaeon *Pyrococcus furiosus*, the crRNAs are further processed to form stable ~35- to 45-nt guide RNAs or prokaryotic silencing RNAs (psiRNAs) (Hale et al., 2008)

The production of crRNAs requires the presence of five Cas proteins - CasABCDE - that reportedly form a complex referred to as Cascade (CRISPR-associated complex for antiviral defense). Of these 5 enzymes, the catalytic activity of the CasE subunit has been shown to be essential for pre-crRNA cleavage, which occurs within each repeat of the CRISPR RNA precursor (Brouns et al., 2008; Carte et al., 2008). Specifically, crRNAs formed in the presence of active CasE 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). Following cleavage, *P. furiosis* Cas6 (a likely CasE functional homolog with similar structure in CasE$^-$ organisms) remains bound to the CRISPR repeat sequences at the 5′ end of the cleavage product (Carte et al., 2008). Whether this enzyme also participates as part of the Cas effector complex, thereby coupling guide RNA biogenesis with interference function, is currently uncertain. Since crRNAs and mature psiRNAs 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 has been suggested previously (Carte et al., 2008; Kunin et al., 2007). The current hypothesis is that the resulting ribonucleoprotein complexes are guided to viral nucleic acids where they silence their genetic target. However, unlike the eukaryotic RNA interference (RNAi) phenomenon to which CRISPR activity was originally compared (Makarova et al., 2006), the target for CRISPR interference seems to be at the DNA level (Brouns et al., 2008; Marraffini and Sontheimer, 2008). At present, the exact mechanism by which Cas effector complexes silence extrachromosomal DNA is not known.

Here, we show that the *E. coli* CRISPR system is triggered by the cell's natural response to envelope stress that resulted from the overexpression of a protein substrate targeted for export by the twin-arginine translocation (Tat) pathway. Activation of CRISPR RNA-mediated interference resulted in specific silencing of the plasmid DNA that encoded the Tat substrate. This silencing required (1) base-complementarity between spacer DNA and the plasmid, (2) specific targeting of the substrate to the Tat export pathway, (3) the BaeSR two-component signal transduction system, and (4) the Cas proteins, in particular a stable ternary complex formed by CasCDE that was observed to function as the catalytic core of the Cas system to process pre-

CRISPR RNA into mature crRNAs. Moreover, all of the Cas proteins except for Cas1 exhibited a certain level of nuclease activity against either ds- or ss-DNA substrates suggesting that a subset of these enzymes have a direct role in the effector function of the CRISPR-Cas system.

2. Results

2.1. Identification of CasE as a key component in prokaryotic gene silencing

In our previous studies of the bacterial Tat export system (Perez-Rodriguez et al., 2007), 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 (Figure 7b and c). We initially suspected that the ssTorA-GFP fusion protein was unstable in the absence of DnaK, rendering it highly susceptible to proteolytic degradation. However, Northern analysis revealed that the absence of ssTorA-GFP in cells was due to elimination of the encoding mRNA (Figure 7d). When plasmid-encoded DnaK was provided *in trans* to Δ*dnaK* cells, the cell fluorescence and *sstorA-gfp* mRNA levels returned to near wild-type levels (Figure 8a and b).

In a search for the factor(s) responsible for silencing the *sstorA-gfp* mRNA, we screened a mini-Tn*10* transposon library in a Δ*dnaK* strain background to isolate suppressor mutants that restored expression, and thus fluorescence, of ssTorA-GFP. As we expected, one of the suppressors mapped to the *yghQ* gene, which is a predicted inner membrane serine

protease. This protease could be responsible for the degradation of misfolded GFP in *⊿dnaK* mutants, whereby the absence of this chaperone can promote its proteolytic activity. But, to our surprise, most suppressors mapped to the *cas* operon and, in particular, to the endoribonuclease encoded by *casE* (previously *ygcH*) (Figure 7a). Because the bioinformatic information on this operon indicates a possible RNAi system (Makarova et al., 2006), which has a growing potential in therapeutics, we decided to focus only on the *cas* operon.

A 'clean' deletion of the *casE* gene in strain BW25113 Δ*dnaK* was created. As expected, cell fluorescence was recovered in the Δ*dnaK* Δ*casE* strain expressing the ssTorA-GFP fusion protein (Figure 7b and c). Restoration of fluorescence in this strain was attributed to stable accumulation of the encoding *sstorA-gfp* mRNA (Figure 7d). To ensure this phenotype was due to inactivation of the *casE* gene 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 (Figure 8a) and accumulated a correspondingly low level of *sstorA-gfp* mRNA (Figure 8b), indicating that CasE was responsible for the gene-silencing phenotype observed in Δ*dnaK* cells expressing ssTorA-GFP. Expression of $CasE^{H20A}$, a His20Ala variant of CasE that is incapable of pre-crRNA cleavage (Brouns et al., 2008), did not result in ssTorA-GFP silencing (Figure 8a). The *E. coli* K12 CRISPR-Cas system comprises an operon of seven *cas* genes: *cas1*, *cas2*, and five genes designated *casABCDE* that encode the Cascade complex (Brouns et al., 2008) (Figure 7a). Individual deletion of each of these genes in a Δ*dnaK* strain background revealed that, in addition to *casE*, all of the other *cas* operon components were also essential for the gene-silencing

activity (Figure 9a). This was corroborated by the absence of ssTorA-GFP silencing in BL21(DE3) Δ*dnaK* cells (Figure 9b), which do not encode a CRISPR-Cas system (Brouns et al., 2008). It should be noted that the absence of DnaK was clearly a requirement for the gene-silencing phenotype because wild-type (wt) MC4100 and single *cas* mutants all emitted strong fluorescence following ssTorA-GFP expression (data not shown).

The *E. coli* K12 genome contains a CRISPR locus that includes three spacers (spacer 5, 8 and 12) that contain sequence complementarity to a region of the *sstorA* gene (Figure 10a). To determine if sequence complementarity between these spacers and the ssTorA signal peptide was responsible for silencing, the plasmid-encoded *sstorA* sequence was modified with synonymous mutations that we predicted would reduce base pairing between the spacer region and the target sequence, but would not change the sequence at the protein level. A total of 2 or 6 synonymous point mutations were introduced in the *sstorA* sequence yielding the constructs ssS8-GFP or ssS5.8.12-GFP, respectively (Figure 10a). The synonymous mutations did not affect targeting as the resulting constructs were expressed and localized in wt *E. coli* cells in a manner that was indistinguishable from unmodified ssTorA-GFP (Figure 10b). Yet when ssS8-GFP and ssS5.8.12-GFP were expressed in Δ*dnaK* cells, the mRNA and fluorescence activity for the synonymous variants was stabilized whereas the mRNA and fluorescence for ssTorA-GFP was silenced (Figure 10c and d). These results indicate that CRISPR interference can silence ssTorA-GFP expression in a manner that is specified by sequence identity between a spacer and a plasmid target sequence.

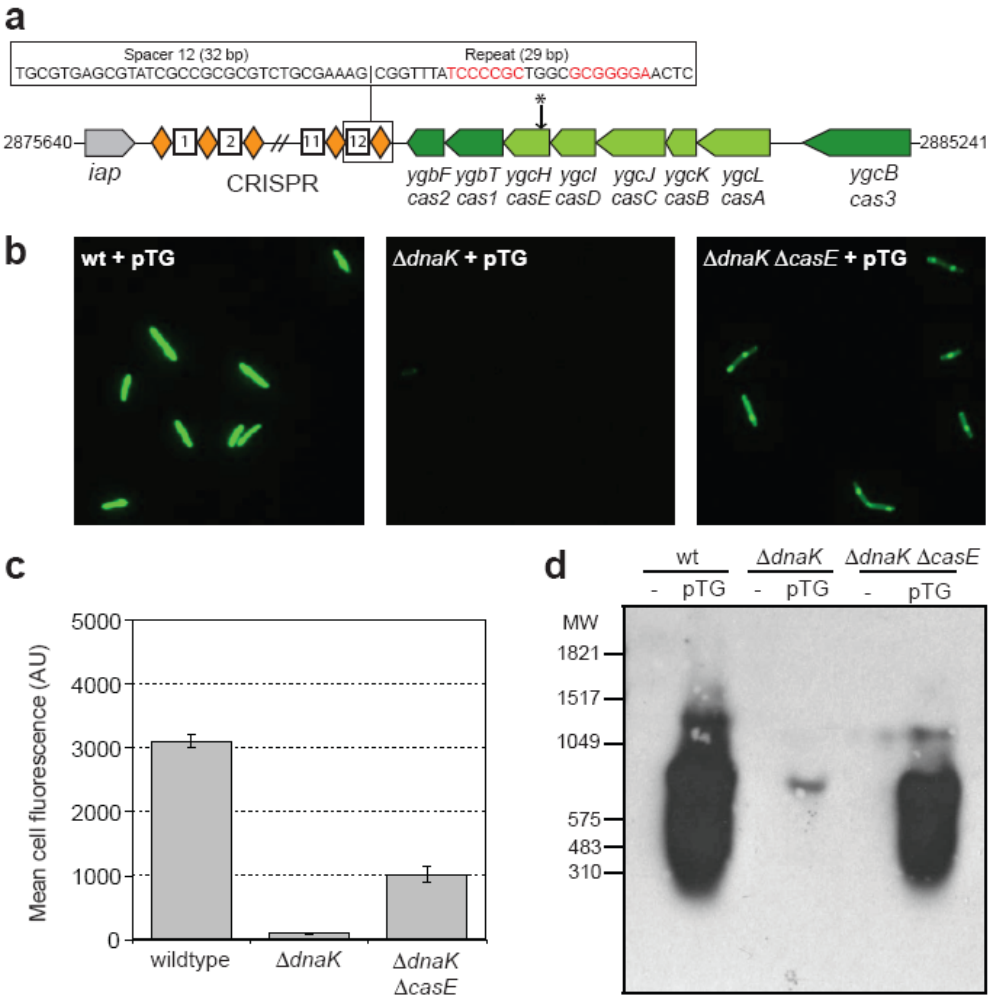


Figure 7. *CasE-dependent silencing of ssTorA-GFP*. (a) The CRISPR-Cas gene cluster of *E. coli* K12. Repeats and spacers are indicated by diamonds and rectangles, respectively. A palindrome in the repeat is indicated by red lettering. Protein family nomenclature is described in (Brouns et al., 2008). Asterisk indicates the relative location of the mini-Tn*10* insertion site in *casE* that was isolated by transposon library screening. (b) Fluorescence microscopy and (c) flow cytometric screening of *E. coli* BW25113 (wt), BW25113 *ΔdnaK* or BW25113 *ΔdnaK ΔcasE* expressing ssTorA-GFP from pTG, as indicated. (d) Northern blot of total RNA isolated from BW25113 (wt), BW25113 *ΔdnaK* or BW25113 *ΔdnaK ΔcasE* cells carrying no plasmid (-) or expressing ssTorA-GFP from pTG. Blot was probed using GFP-specific digoxigenin-labeled oligonucleotides.

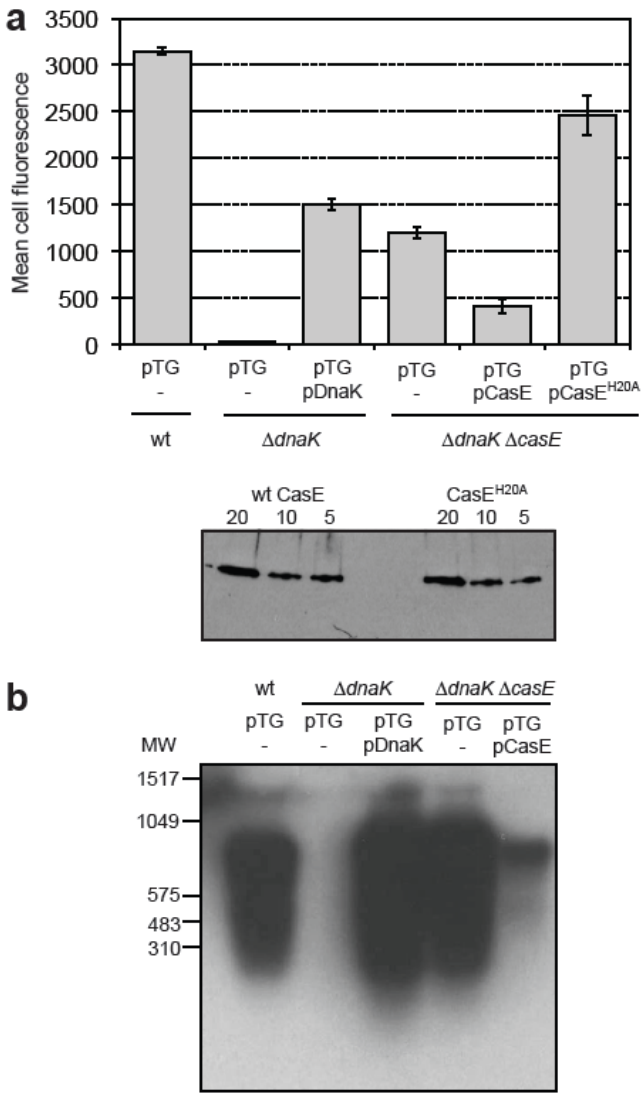


Figure 8. *Genetic complementation of the silencing phenotype. (a*) Upper panel, flow cytometric analysis of BW25113 wt, *ΔdnaK* and *ΔdnaK ΔcasE* cells expressing ssTorA-GFP and co-expressing DnaK, CasE or CasE^{H20A} as indicated. Lower panel, Western blot analysis of wt CasE and CasE^{H20A} following purification from an equivalent number of cells. Varying volumes (20, 10 and 5 ml) were loaded as indicated. Blot was probed with anti-polyhistidine antibodies. (b) Northern blot analysis of total RNA isolated from the same cells as in (a). Blot was probed using GFP-specific digoxigenin-labeled oligonucleotides.

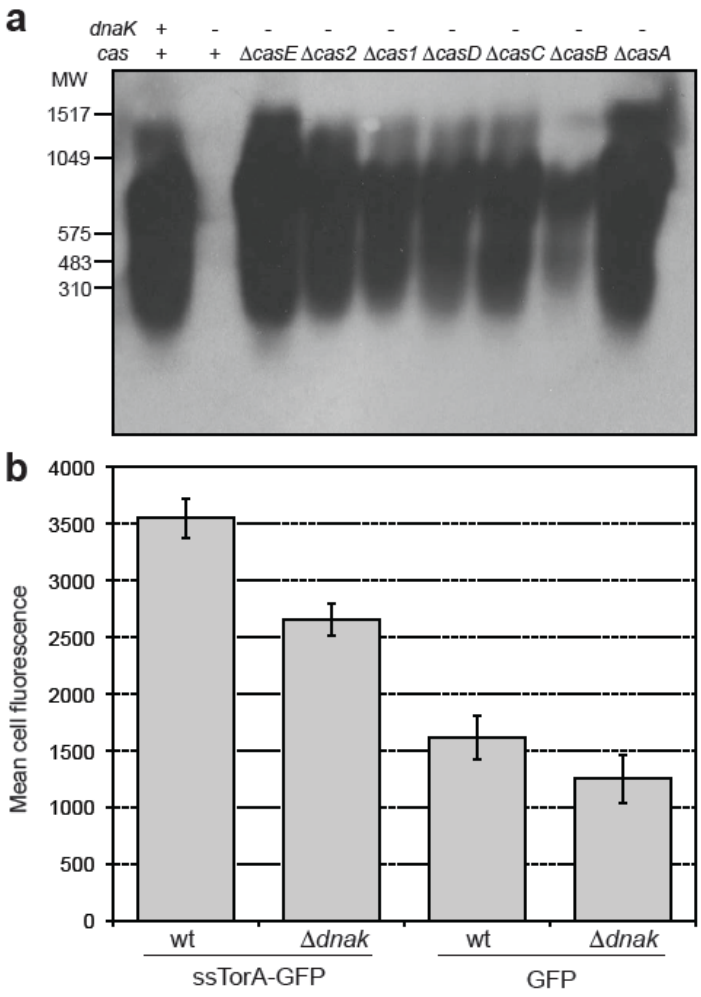


Figure 9. *Contribution of the Cas system to ssTorA-GFP silencing. (a)* Northern blot analysis of total RNA isolated from BW25113 wt (lane 1), *ΔdnaK* (lane 2) and *ΔdnaK ΔcasX* (lanes 3-9, where *X* = different *cas* genes as indicated) expressing ssTorA-GFP. Blot was probed using GFP-specific digoxigenin-labeled oligonucleotides. (b) Flow cytometric analysis of BL21(DE3) wt and *ΔdnaK* cells expressing ssTorA-GFP or GFP as indicated.

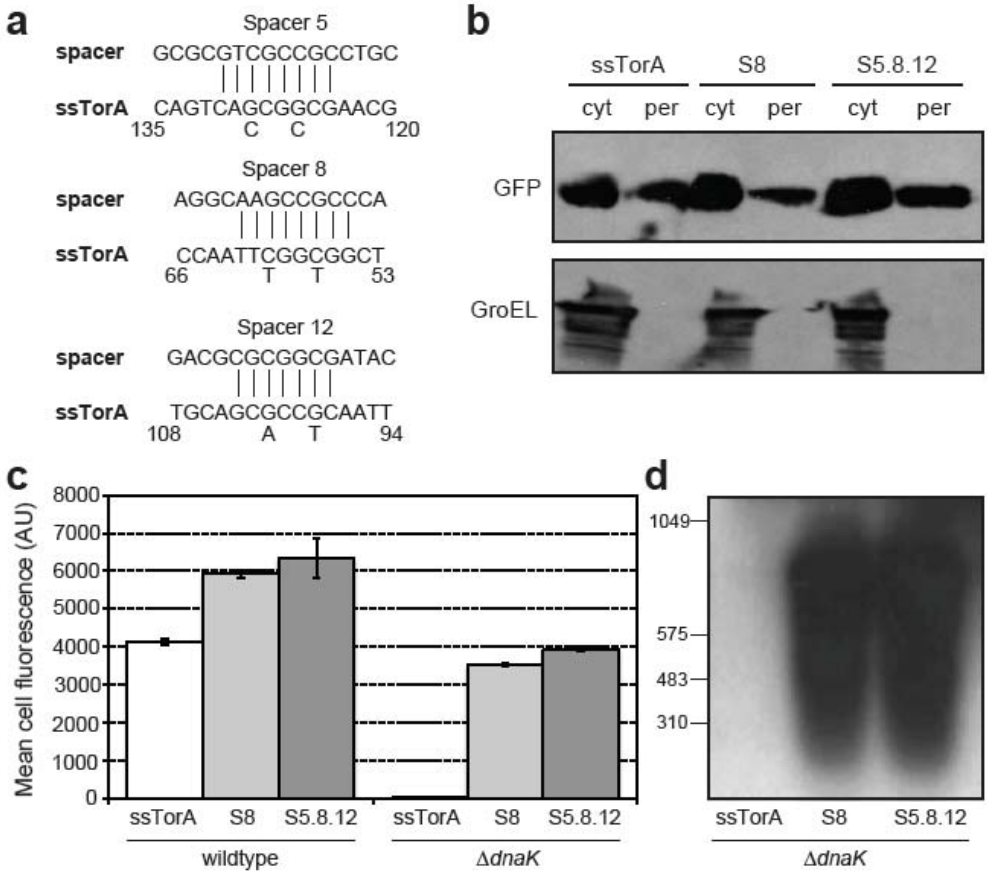


Figure 10. *Complementarity between spacer DNA and ssTorA sequence is required for silencing activity.* (a) *E. coli* K12 spacers 5, 8 and 12 aligned with the sense strand encoding the ssTorA signal peptide. Numbering of ssTorA corresponds to the bp position relative to the start codon of native *E. coli* TorA. Synonymous base changes are indicated below the ssTorA sequence. The pssS8-GFP construct was created by mutating the C57 and C60 nucleotides in *sstorA* to T57 and T60, respectively. The ss5.8.12-GFP construct was created by further mutating ssS8-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, ssS8-GFP or ssS5.8.12-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 *ΔdnaK* cells expressing ssTorA-GFP, ssS8-GFP or ssS5.8.12-GFP. (d) Northern blot of total RNA from BW25113 *ΔdnaK* cells expressing ssTorA-GFP, ssS8-GFP or ssS5.8.12-GFP. Blot was probed using GFP-specific digoxigenin-labeled oligonucleotides.

2.2. Reconstitution of the core Cas complex

Having confirmed that ssTorA-GFP silencing involves genomic spacers and Cas proteins, we next sought to understand how the Cas enzymes process the pre-crRNAs into crRNAs, a key step in the initiation of CRISPR silencing. Since all of the *cas* genes tested were required for silencing activity (Figure 9a) and since the CasABCDE proteins were previously shown to co-purify (Brouns et al., 2008), we attempted to reconstitute the *E. coli* Cas complex in a series of mixing and co-expression experiments (Figure 11a and b). The CasABCDE proteins as well as Cas1, Cas2, and Cas3 were purified individually. While many of these proteins behaved well on their own in solution, CasC oligomerized into a ~490-kDa species (Figure 11a), and CasE interacted strongly with nucleic acids. CasD was only soluble when in complex with CasC, the resulting complex migrated as a 105-kDa species on the sizing column (Figure 11a). We estimated a molar ratio of 2:1 for the CasCD complex based on their Coomassie-stained band intensities on the SDS gel (Figure 11c). Strong interactions were also detected between CasC and CasE, which form a ~1:1 molar ratio complex migrating at roughly 832-kDa on the sizing column (Figure 11a and d). Two different reconstitution schemes then confirmed that these three Cas proteins further form a salt-stable ternary complex (Figure 11d). This complex is around 200 kDa when CasD is fused with an N-terminal SUMO tag (Figure 11a). Solubility is reduced in the absence of the SUMO tag, and the ternary complex migrates as a >600 kDa species with an apparent stoichiometry estimated at 2:1:2 for the CasC, CasD, and CasE, respectively (data not shown). CasC appears to be the center of the complex since no interactions can be found between CasD and CasE (data not shown). Previously, CasA and CasB were found at

substoichiometric levels in the affinity purified *E. coli* Cascade complex (Brouns et al., 2008), however in our hands incubating CasA and B with the CasCDE complex did not lead to higher complex formation under stringent or 'physiological' conditions (data not shown) which led us to conclude that they are not stable components of the core Cas complex. Consistent with this view, we show below that CasA and CasB do not alter the pre-crRNA processing pattern by the CasCDE ternary complex.

To investigate whether the molecular composition of the Cascade complex is conserved phylogenetically, we carried out similar reconstitution of the *T. thermophilus* Cas system. Our results revealed an identical ternary complex between the *T. thermophilus* CasCDE proteins (data not shown). Furthermore, we demonstrated that *T. thermophilus* CasE, which is ~35% identical to *E. coli* CasE, could hetero-assemble with *E. coli* CasC or CasCD to form salt-stable higher complexes (data not shown). We therefore expect that CasCDE-containing bacteria form a conserved Cas core complex.

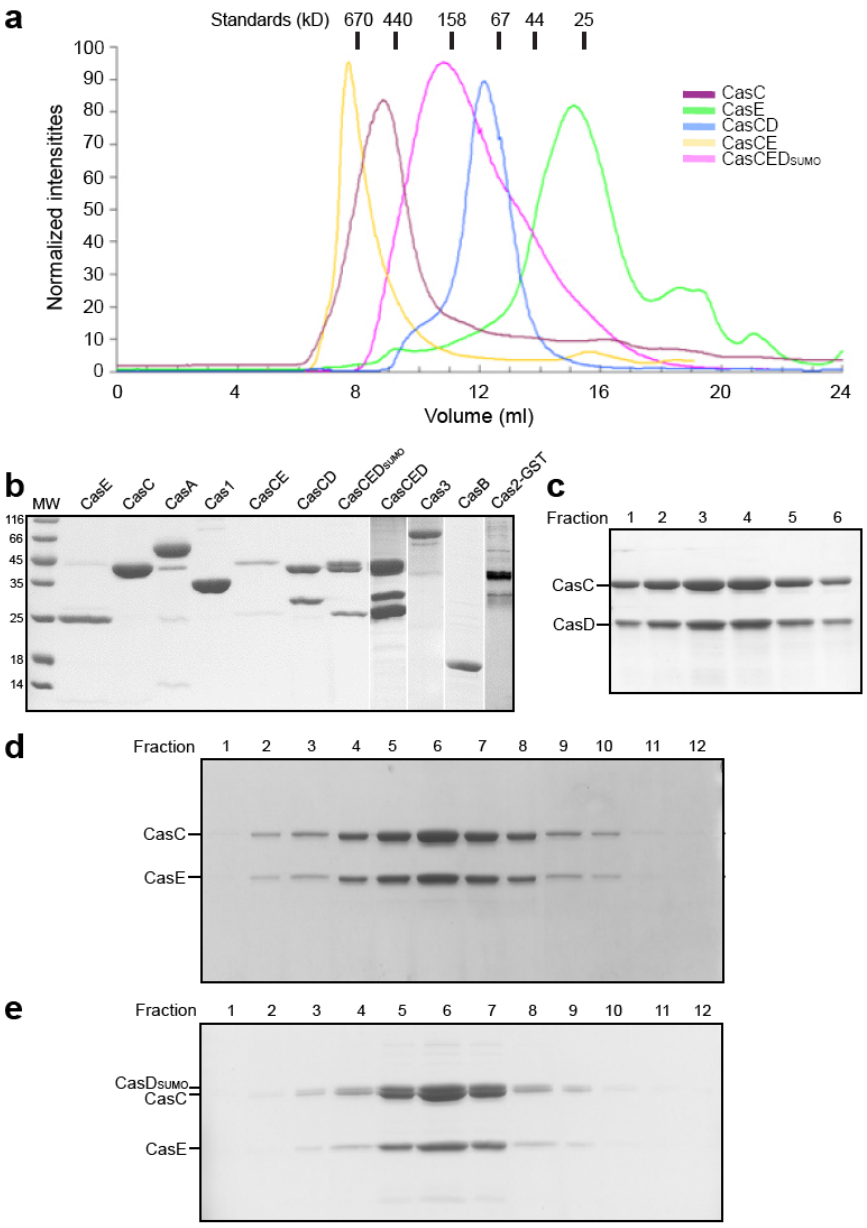


Figure 11. *Reconstitution of E. coli Cascade core complexes. (*a) Elution profile of CasC and CasE proteins and CasCD, CasCE, and CasCDE complexes on Superdex 200 size exclusion column. (b) SDS-PAGE analysis of the individual Cas proteins and Cas complexes. (c-e) SDS-PAGE analysis of the size exclusion fractions of the (c) CasCD, (d) CasCE, and (e) $CasCED_{SUMO}$ complexes.

2.3. CRISPR RNA processing by core Cas complex

To reveal the function of the Cas proteins in CRISPR RNA processing, we next carried out *in vitro* RNA processing assays using either individual components or the Cas core complexes. Previous studies showed that processing of pre-crRNA into unit length in *E. coli* requires the participation of Cas proteins, especially a catalytically competent CasE protein (Brouns et al., 2008; Carte et al., 2008). CasE-mediated processing mainly involves two endonucleolytic events in the CRISPR repeat region to release the spacer RNA: an invariable cleavage site at the base of the 5' hairpin structure and a variable cleavage site at 3-4 different positions in the vicinity of the hairpin loop (Brouns et al., 2008; Carte et al., 2008). Here, two 5'-^{32}P or fluorescently labeled CRISPR RNA substrates were used, one encoding the entire spacer 5' region flanked on each end by a complete CRISPR repeat, the other containing just the conserved CRISPR repeat (Figure 12a). When individually purified Cas components were incubated with a 5'-HEX-labeled CRISPR spacer 5 sequence, only Cas2, CasC and CasE showed strong RNase activity (Figure 13). The previously reported catalytic mutant CasE^{H20A} still retained certain catalytic activity, but was much more promiscuous than wt CasE (Figure 13). The same reactivity pattern was observed when the CRISPR repeat or CRISPR-spacer5-CRISPR RNAs were used as substrates (data not shown). We then carried out a more thorough analysis of the cleavage products produced by CasC, CasE, CasCE, and the CasCDE (in which CasD is N-terminally tagged with SUMO) complexes on sequencing gels. Interestingly, neither CasC nor CasE cleaved the two CRISPR RNA substrates at the cleavage sites reported previously (Brouns et al., 2008). Rather, CasE cleaved in the spacer region and in the general vicinity of the CRISPR loop

region, and its cleavage pattern was similar to that of RNase A (Figure 12b), whereas CasC displayed a rather distributive cleavage pattern (Figure 12b). The enzymatic behavior of CasE is in sharp contrast with its functional homolog Cas6 from archaeal *P. furiosus* (Carte et al., 2008), which single-handedly makes a distinct cut in the hairpin loop of the CRISPR repeat. This may be an indication that the archaeal organisms have a distinct mechanism for pre-crRNA processing, since components of the bacterial Cascade complex, such as CasE, are often not found in the archaeal species (Brouns et al., 2008). The nonspecific cleavage pattern was replaced with a new, specific cleavage pattern following incubation with the CasCE complex. Specifically, CasCE cleaved the CRISPR repeat substrate at predominantly two locations within the hairpin (named CS1 and CS2). By comparison with the alkaline hydrolysis ladder, we determined that CS1 coincides with the invariable cleavage site that defines the 5'-end of the processed crRNA (Brouns et al., 2008), whereas CS2, which is rather promiscuous and is predominantly comprised of two sites adjacent to each other, is two nucleotides away from the most popular 3'-processing site inferred from the CRISPR RNA cloning results (Brouns et al., 2008). Interestingly, this site-specific cleavage pattern was further modulated following incubation with the ternary Cas core complex, as CS2 now shifts upstream and expands to three adjacent nucleotides, overlapping precisely with the most popular 3'-processing site (Brouns et al., 2008). In the case of the natural CRISPR-spacer 5-CRISPR RNA, the nonspecific cleavage events in the spacer region disappeared and the same specific cleavage pattern emerged when the substrate was incubated with the same set of Cas proteins and complexes (Figure 12b), suggesting that the Cas core complex specifically targets the repeat region, but not the spacer

region. It is noteworthy that addition of CasA, CasB, CasA/B, or Cas1/3 proteins together with CasCDE did not appreciably change the CRISPR processing pattern or rate (Figure 12c). We also verified that the cleavage patterns of the SUMO-tagged and untagged CasCDE complexes were identical in our assay conditions (Figure 12c), eliminating the possibility that the N-terminal SUMO tag in CasD altered the biochemical behavior of the CasCDE complex. All together, we demonstrated that the CasCDE core complex is necessary and sufficient for the correct processing of pre-crRNAs.

To definitively assign the cleavage activities inside the Cas core complex, we next characterized the biochemical behavior of $CasE^{H20A}$ by itself and in complex with other Cas proteins. As stated above, H20A was previously shown to completely abolish the production of crRNA (Brouns et al., 2008). In our RNase assay, $CasE^{H20A}$ still possessed weak, rather nonspecific RNase activity (Figure 12c). Comparison of the cleavage pattern of the $CasE^{H20A}$ containing complexes with their wt counterparts revealed that the distinct CS1 cleavage site became promiscuous in the mutants, while the cleavage pattern in the CS2 region was identical (compare Figure 12c with Figure 12b). Assuming the altered CS1 site was due to the CasE mutation, our data are consistent with a mechanistic model in which CasE is responsible for CS1 cleavage and CasC is responsible for CS2 cleavage. The substrate specificity is achieved through the formation of a stable CasCDE ternary structure. Our data did not, however, completely rule out the less likely possibility that CasC is responsible for both CS1 and CS2 cleavage, and that the disappearance of CS1 is due to an alteration of ternary structure. Identification and characterization of CasC mutants will provide definitive answers to resolve this uncertainty.

Figure 12. *RNase activity of Cas proteins and CasE-containing complexes.* (a) Two RNA substrates used in the activity assay. Left, CRISPR-Spacer5-CRISPR contains the natural *E. coli* Spacer 5 sequence flanked by two CRISPR repeats. Right, the conserved *E. coli* CRISPR repeat sequence. The cleavage sites by CasCE and CasCDE core were marked by lines, and their relative intensity was of the order (highest to lowest): CS1 ≈ 3 > 2 > 4 > 1. (b) Cleavage profile of CasE, CasC, CasCE, and $CasCED_{SUMO}$ core complexes resolved on 16% sequencing gel. 5'-^{32}P-labeled CRISPR-Spacer5-CRISPR was the substrate used in the left half of the gel; 5'-^{32}P-labeled CRISPR was used in the right half. RNA alone, RNase A partial digestion, and alkaline hydrolysis ladder lanes were included for each substrate as controls. RNAs were gel-purified after labeling reaction. The amount of protein used in the assay was from left to right 0.125, 1.25, and 12.5 μM. An empty lane was removed during figure generation. The cleavage site was assigned with reference to the hydrolysis ladder, which is best resolved in panel (c). $CasE^{H20A}$ mutant was assayed using the CRISPR repeat RNA (Figure 12a, right) that was 5'-^{32}P-labeled and used directly after removing unincorporated nucleotides. Alkaline hydrolysis ladder in lane 1 resolves the position of every base on the 16% sequencing gel. The amount of protein used in the assay was from left to right 0.125, 1.25, and 12.5 μM. Removing the N-terminal SUMO tag on CasD did not change the cleavage profile of CasCDE. Included in the four right lanes is ~5 μM of either CasA, CasB, CasA+CasB, or Cas1+Cas3 proteins together with 1.25 μM of CasCDE in the cleavage reaction.

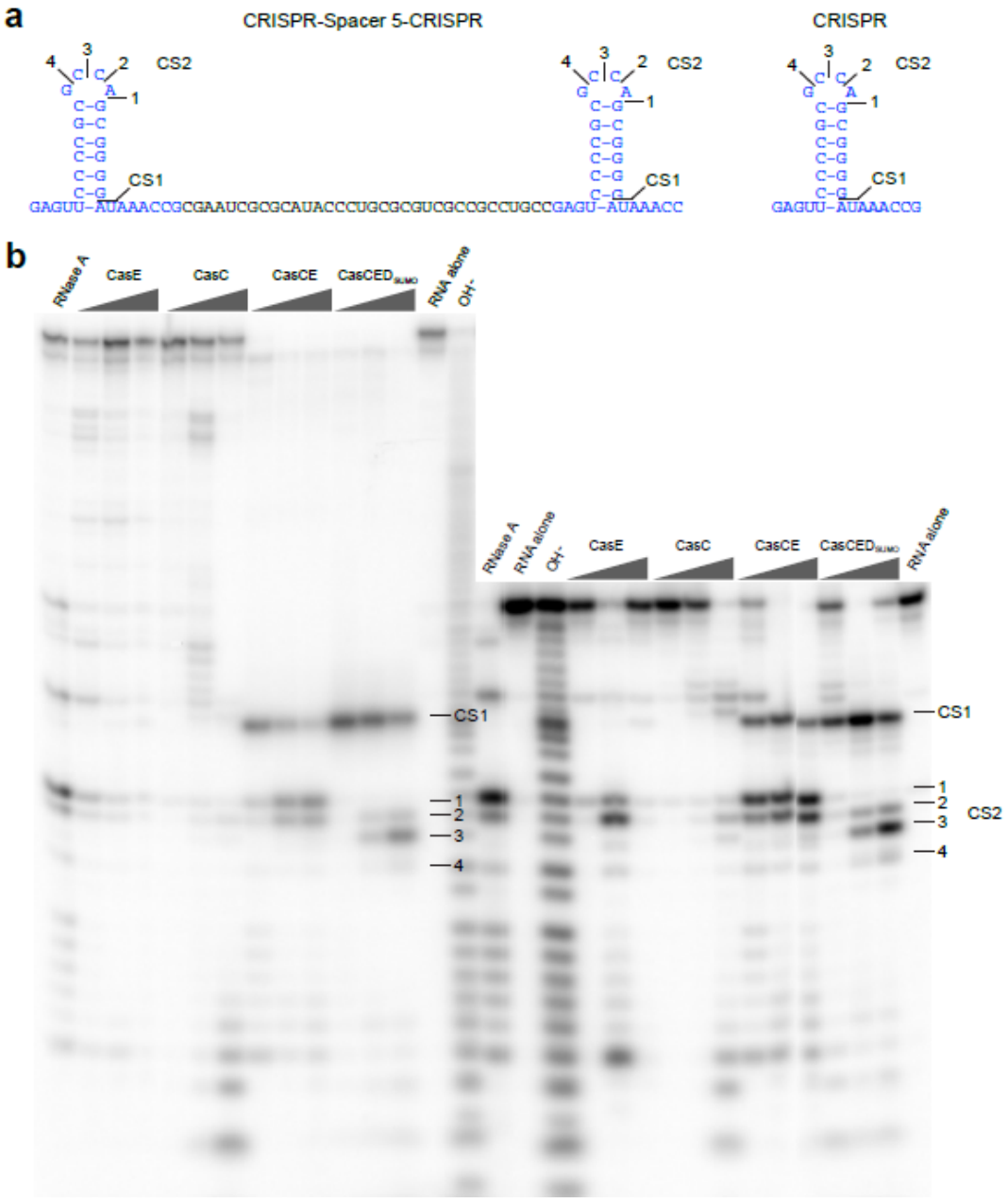
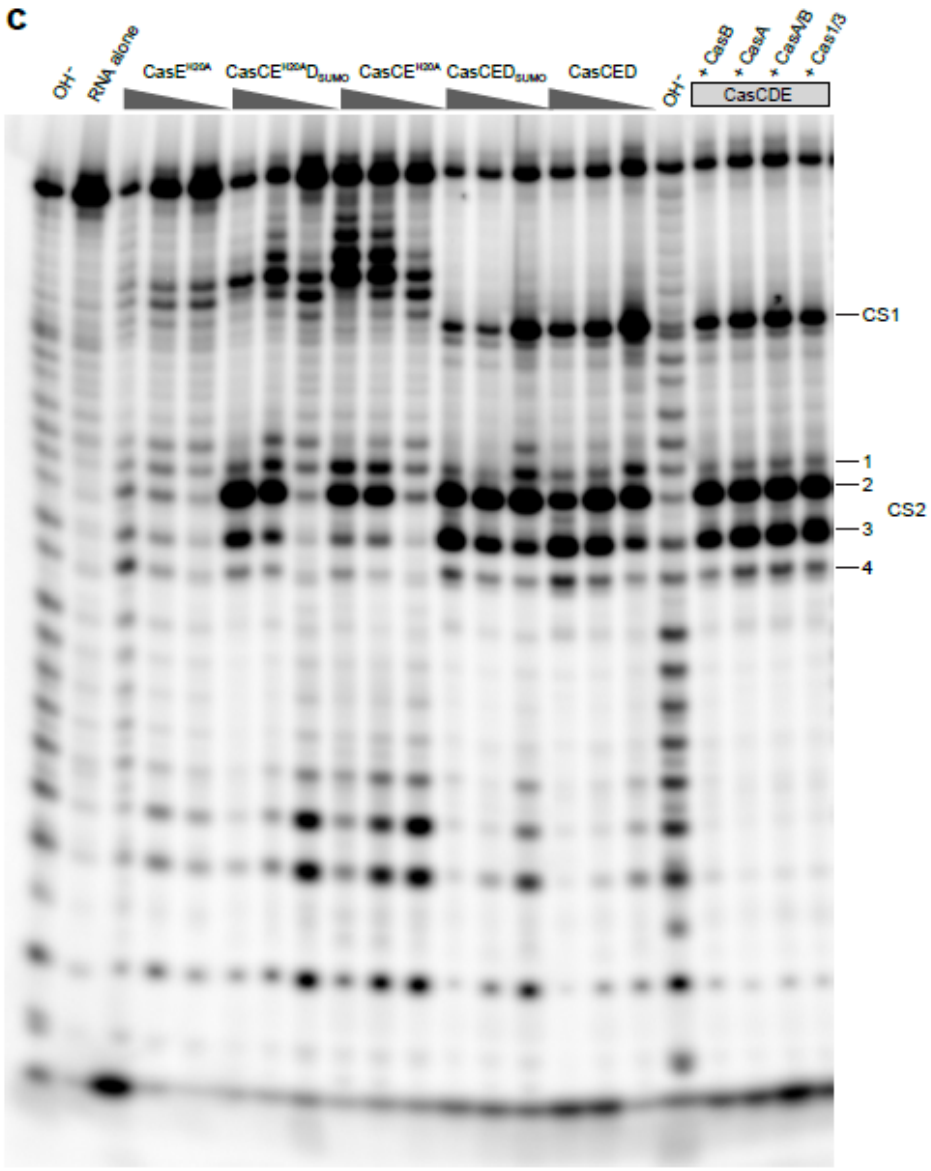

a
CRISPR-Spacer 5-CRISPR
CRISPR
CS2
CS1
GAGUU-AUAAACCGCGAAUCGCGCAUACCCUGCGCGUCGCCGCCUGCCGAGU-AUAAACC
GAGUU-AUAAACCG
b
RNase A
CasE
CasC
CasCE
CasCED$_{SUMO}$
RNA alone
OH-
CS1
CS2
1
2
3
4

Figure 12 (Continued)



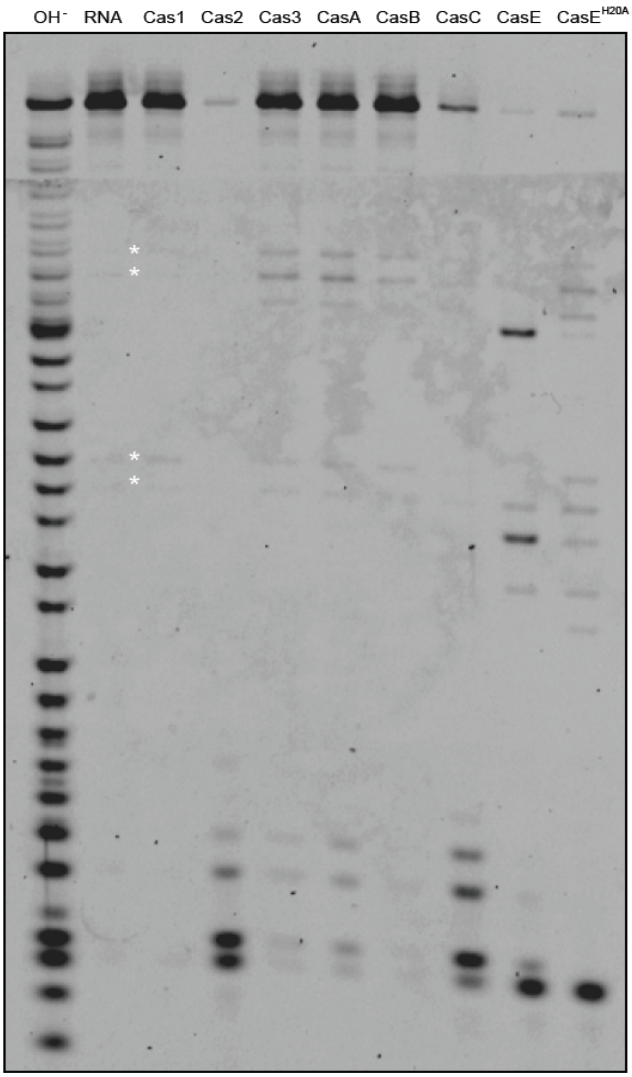


Figure 13. *Survey of RNase activity among Cas proteins.* Each purified Cas protein (~1 μM) was incubated with 50 nM of chemically synthesized 5'-HEX-labeled CRISPR repeat RNA (see Figure 12a). Alkaline hydrolysis ladder and RNA alone are shown in lanes 1 and 2, respectively. The background bands irrelevant to enzymatic cleavage are highlighted by white asterisks. We use the appearance of strong endonucleolytic cleavage bands and the accumulation of 3-7-bp products as an indication of robust RNase activity. The bands in Cas3, CasA, and CasB are only slightly stronger than the overlapping background, and may have come from minute RNase contamination.

2.4. Envelope stress triggers Cas-mediated gene silencing via BaeSR two-component signal transduction system

Since the Cas proteins characterized above were encoded in the genomes of both wt and *ΔdnaK* cells, we were surprised that the gene-silencing phenotype was only observed in *ΔdnaK* cells and not in wt cells. To address this issue, we explored the factors that were essential for the silencing activity in *ΔdnaK* cells. Previously, we observed that GFP expressed without an N-terminal Tat signal peptide accumulated to very high levels and was active in the cytoplasm of cells that lacked a functional copy of DnaK (Perez-Rodriguez et al., 2007). This suggested that targeting of GFP to the Tat translocase in the absence of DnaK might be a trigger for Cas-mediated silencing. Indeed, when the twin arginines in the Tat signal peptide of ssTorA-GFP were replaced with twin lysine residues (TorA(R11K:R12K)), amino acid substitutions that completely block Tat-dependent GFP export (DeLisa et al., 2002), no silencing of ssTorA-GFP was observed in *ΔdnaK* cells (Figure 14a and b). Likewise, there was also no measurable silencing when the export-competent ssTorA-GFP fusion was expressed in *ΔdnaK* cells that also lacked the essential Tat translocase component TatC (Fig 14a and b).

Based on these findings, we hypothesized that membrane localization of the highly expressed ssTorA-GFP fusion in *ΔdnaK* cells might induce an envelope stress, perhaps due to membrane insertion of GFP that is misfolded in the absence of DnaK chaperone activity. In response to this stress, 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 is comprised of the membrane-localized histidine kinase BaeS that senses the envelope stress and accordingly modulates the phosphorylation state of the cytoplasmic transcription factor BaeR (MacRitchie et al., 2008). We focused our attention on this signaling 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 activity in *⊿dnaK* cells that also lacked BaeS or BaeR (Figure 14c).

Since BaeSR was clearly 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* (Figure 14d) 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* (Figure 14d). This confirms that complexes formed by BaeR-His_6 and the BaeR-box sequence internal to *casA* detected in this experiment were specific.

2.5. Cas-mediated ssTorA-GFP silencing occurs via targeting of plasmid DNA

As mentioned above, CRISPR interference acts at the DNA level and therefore is fundamentally different from eukaryotic RNAi. In light of this, we next investigated whether DNA molecules encoding *sstorA-gfp* were targeted by the Cas machinery in *ΔdnaK* cells. First, we isolated plasmid pTG from an equivalent number of wt, *ΔdnaK* and *ΔdnaK ΔcasE* cells following the induction of ssTorA-GFP and performed restriction analysis on the resulting plasmid DNA preparations. The pTG plasmid was stably maintained in wt and *ΔdnaK ΔcasE* cells but was not detected in preparations from *ΔdnaK* cells (Figure 15a). The silencing of pTG in *ΔdnaK* cells was dependent on the *sstorA-gfp* sequence because empty pTrc99A plasmid DNA (the backbone in which *sstorA-gfp* was cloned) was observed to accumulate at a similar level in all strain backgrounds including *ΔdnaK* cells (Figure 15a). Second, we reasoned that these differences in plasmid DNA levels would cause measurable differences in antibiotic resistance conferred by the *ampR* gene encoded elsewhere on plasmid pTG. Indeed, the minimum inhibitory concentration (MIC) on ampicillin (Amp) of wt and *ΔdnaK ΔcasE* cells was 8- and 6-fold greater than that for *ΔdnaK* cells (Figure 15b). Finally, to verify that

the Cas machinery only targets plasmid DNA and not "self" DNA, we determined the viability of wt, *ΔdnaK* and *ΔdnaK ΔcasE* cells by plating cells on non-selective medium following the induction of ssTorA-GFP. The number of colony-forming units (CFU) of *ΔdnaK* expressing ssTorA-GFP was actually slightly greater than that measured for wt and *ΔdnaK ΔcasE* cells (Figure 15c), indicating that the Cas effector complexes specifically target extrachromosomal DNA and avoid targeting genomic DNA.

2.6. General DNA cleavage and nicking activities in Cas proteins

Our observation above that plasmid DNA is the target of the CRISPR-Cas system is consistent with other recent studies (Brouns et al., 2008; Marraffini and Sontheimer, 2008). To test whether the Cas complex itself may be involved in the plasmid degradation observed above, we measured the DNase activity of individual Cas components and the Cas core complex against linear, nicked, and supercoiled ds-plasmid DNA. Under low enzyme:substrate ratio conditions (0.2 μg protein : 1 μg plasmid), CasE displayed the strongest DNA cleavage activity, it also dominated the behavior of the CasCDE complex (Figure 15d). CasA, CasB, and CasC appeared to possess nicking rather than cleavage activity, as the supercoiled DNA was preferentially converted to the nicked form, then to the linear form (Figure 15d). When the same plasmid DNA was linearized first by restriction enzyme *Bam*HI, only CasE and CasE-containing complexes, but not CasA and CasC, displayed significant cleavage activity, evidenced by the downward smearing of DNA, consistent with CasA and CasC possessing nicking activity, which is masked by extensive base pairing (Figure 15d). In parallel, a set of experiments were performed under high enzyme:substrate ratio conditions

(1 μg protein : 0.1 μg plasmid), in which CasA and CasE displayed the strongest activity, followed by CasC, and to a lesser extent CasB (Figure 16a). Under such conditions, CasB and CasC still preferentially degraded the supercoiled DNA (Figure 16a), while CasA was able to degrade all three forms of plasmid as efficiently as CasE. The activity of CasE appeared to be down-regulated upon forming higher order complexes with CasC and CasD (Figure 16a). The H20A mutation had a similar inhibitory effect on the DNase activity of CasE as on its RNase activity, which is consistent with the presence of a single active site for both activities. None of the endodeoxyribonuclease activities are site specific, as using pTG encoding ssTorA-GFP or an unrelated pUC19 plasmid did not alter the experimental outcome (data not shown). Besides the ds-DNase activity, we also tested the activities of the Cas proteins against 5'-Cy5-labeled ss-DNA substrates derived from either the sense or anti-sense strand of spacer 5. The results showed that in addition to CasA, CasB, CasC, CasE, and CasE-containing complexes, ss-DNase activities were also detected for Cas2 and Cas3, leaving Cas1 the only inactive Cas factor (Figure 16b). Cas2 was previously characterized as a sequence-specific endoribonuclease adopting a ferredoxin-like fold (Beloglazova et al., 2008). In our hands, it behaved as an efficient processive RNase and a nonspecific ss-DNase, but not a ds-DNase. Cas3 is a putative DEAD-box helicase often fused to a HD nuclease domain. We also conclude that the HD nuclease in Cas3 is specific for ss-DNA, and possesses background activity against RNA or ds-RNA.

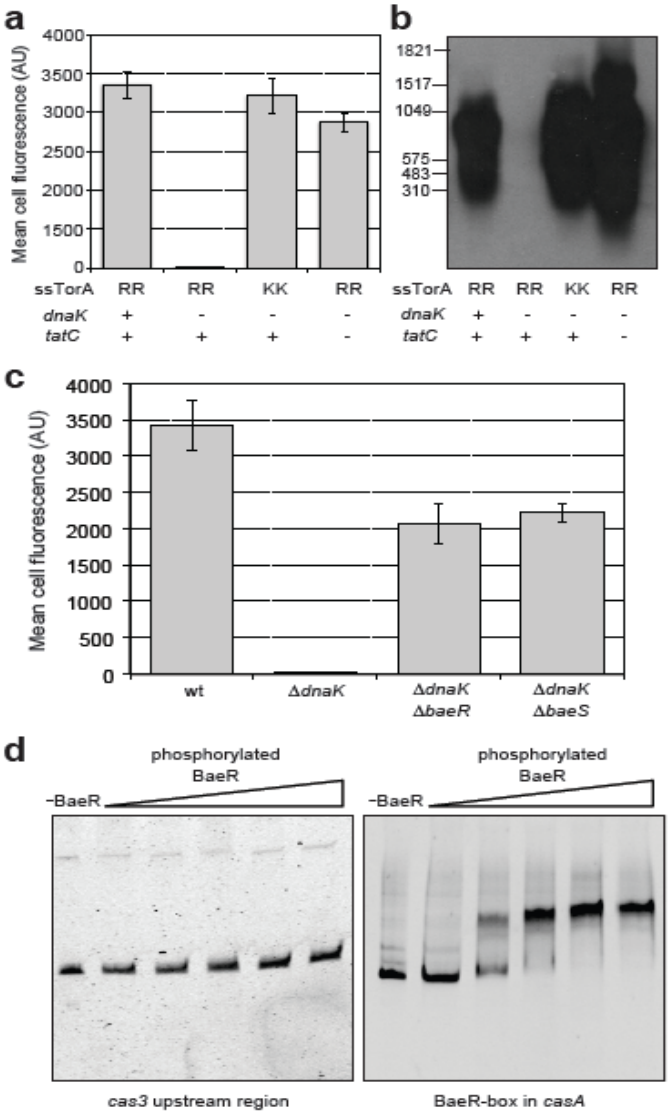


Figure 14. *CRISPR-Cas silencing is dependent on membrane targeting and the BaeSR two-component stress response system.* (a) Flow cytometric screening of BW25113 wt, *ΔdnaK* and *ΔdnaK ΔtatC* cells expressing either ssTorA-GFP or ssTorA(KK)-GFP as indicated. (b) Northern blot analysis of total RNA from the same cells as in (a). Blot was probed using GFP-specific digoxigenin-labeled oligonucleotides. (c) Flow cytometric screening of BW25113 wt, *ΔdnaK*, *ΔdnaK ΔbaeS* and *ΔdnaK ΔbaeR* 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.

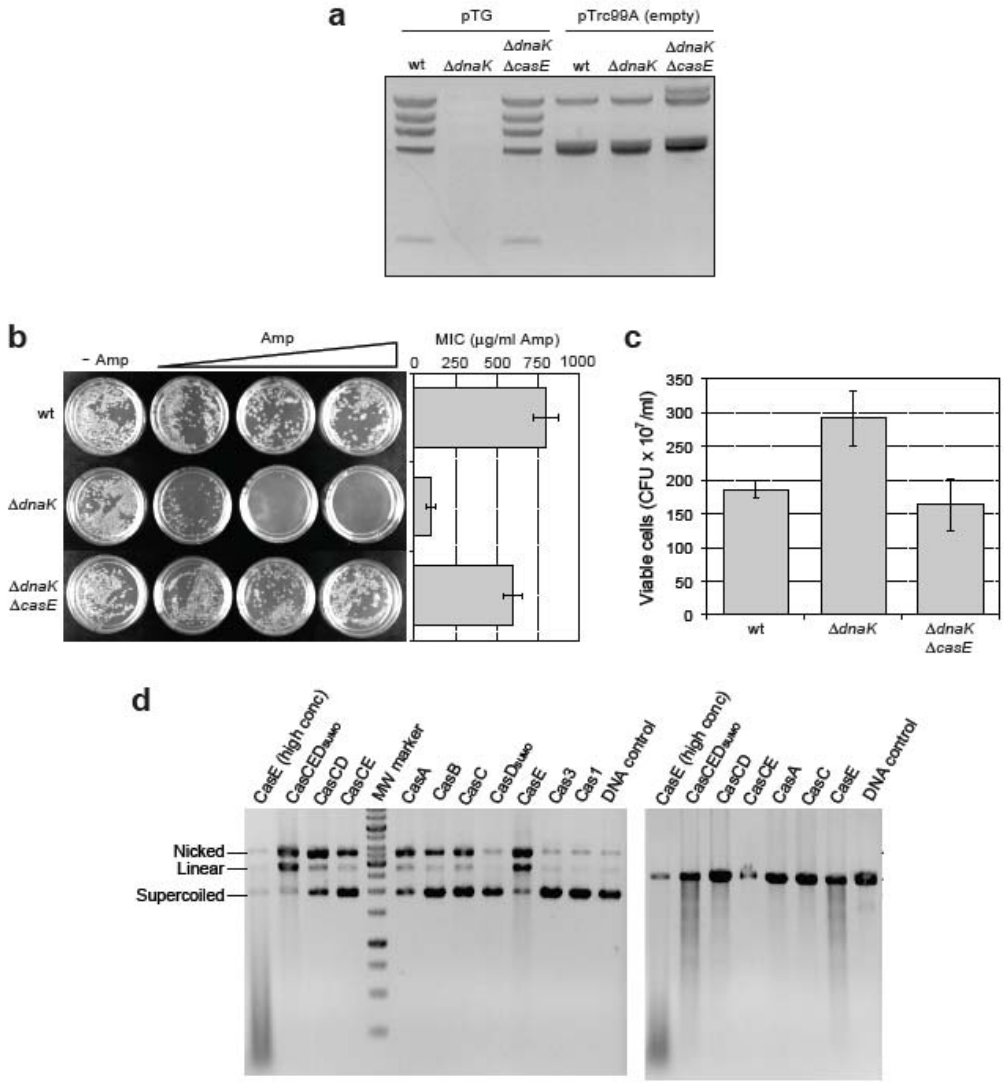


Figure 15. *Plasmid DNA is the target of CRISPR-Cas interference***.** (a) *Rsa*I restriction analysis of plasmid DNA isolated from BW25113 wt, *ΔdnaK* and *ΔdnaK ΔcasE* cells transformed with either pTG expressing ssTorA-GFP or empty pTrc99A (control) as indicated. (b) Selective plating of BW25113 wt, *ΔdnaK* and *ΔdnaK ΔcasE* 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 Amp. (c) Viability of the same cells as in (b). The number of CFUs was determined by plating an equivalent number of induced cells on LB agar with no antibiotics. (d) Survey of double-stranded DNase activity among Cas proteins. Purified Cas proteins and higher complexes were assayed for their activity on either supercoiled (left panel) or linearized (right panel) plasmid under low enzyme:substrate condition (0.2 μg protein : 1 μg plasmid).

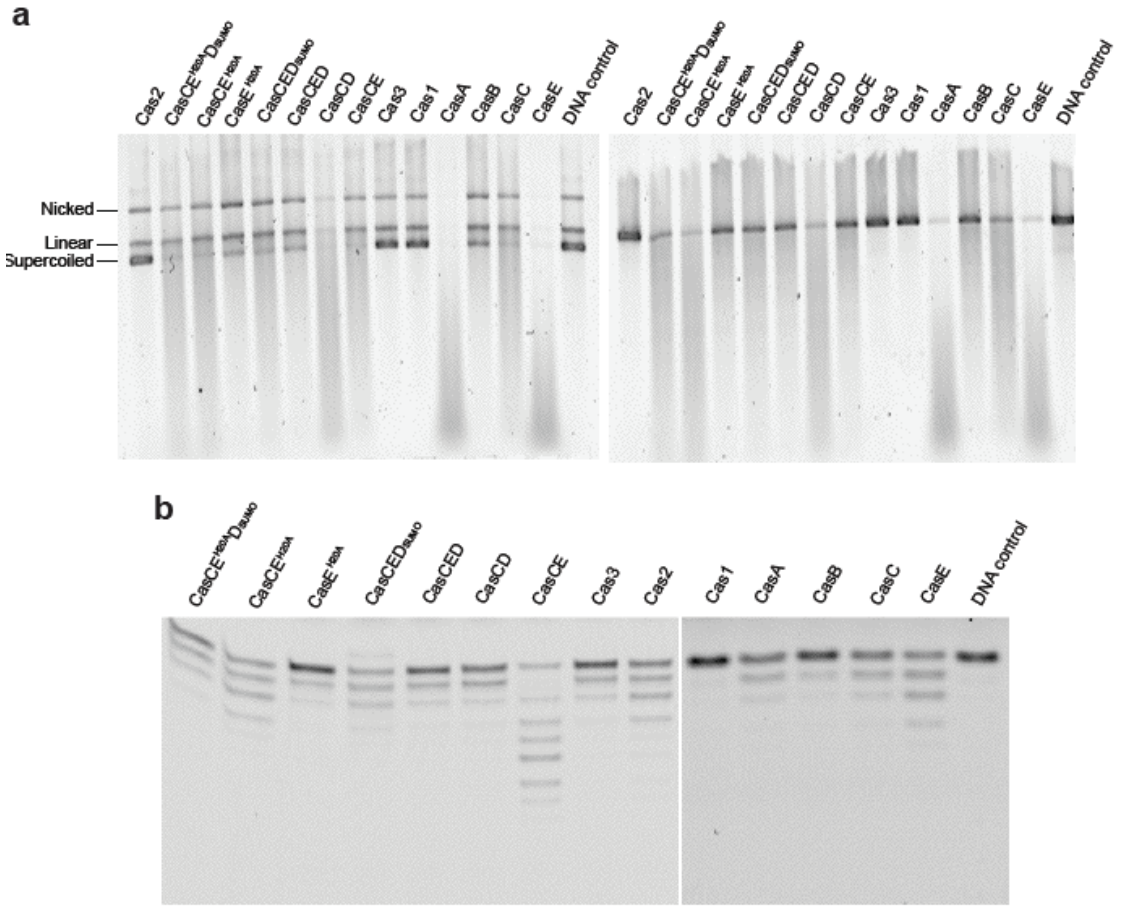


Figure 16. *Survey of ds- and ss-DNase activity under high enzyme:substrate conditions.* (a) Each of the purified Cas proteins and higher complexes were assayed for their activity on either the supercoiled plasmid (left panel) or the linearized form (right panel) under high enzyme:substrate condition (1 μg protein : 0.1 μg plasmid). After 1 hr incubation, the reactions were quenched by phenol extraction and the resulting DNA was separated on a 0.8% agarose gel, then stained by a sensitive Nancy-520 dye (Sigma) and scanned by Typhon 9200 for visualization. (b) 100 nM of a 26-bp 5'-Cy5-labeled DNA oligo was incubated with ~2 μM Cas proteins and higher complexes to reveal their ss-DNase activity. After 1 hr incubation, the reactions were quenched by phenol extraction and the resulting DNA was separated on a 19% PAGE gel and scanned by Typhon 9200 to resolve the cleavage products.

3. Discussion

RNAs can modulate a wide variety of biological processes. In the present report, we demonstrate that a unique class of recently discovered regulatory RNAs known as CRISPR RNAs, together with the Cas proteins, are responsible for silencing extrachromosomal DNA in response to envelope stress. CRISPR RNA-mediated silencing of foreign DNA can be roughly separated into four distinct stages: (1) transcription of pre-crRNA and Cas proteins triggered by cellular cues; (2) processing of pre-crRNA into mature crRNAs; (3) sequence specific targeting of extrachromosomal DNAs for degradation; and (4) acquired immunity through integration of short extrachromosomal DNAs in the CRISPR DNA array of the host genome. In this study, we provide mechanistic insights into the first two stages of this pathway, and shed light on the poorly understood targeting stage (Figure 17).

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 CRISPR-Cas also protects bacteria from protein localization defects that threaten the integrity of the inner cytoplasmic membrane. Specifically, membrane localization of ssTorA-GFP in *ΔdnaK* cells induced an envelope stress that activated the Cas proteins and led to silencing of the ssTorA-GFP-encoding plasmid. Silencing activity was dependent upon the BaeSR signal transduction pathway, indicating for the first time that an endogenous cellular mechanism is involved in triggering the CRISPR-Cas defense system. At present, it is not known why ssTorA-GFP in *ΔdnaK* cells elicited a stress

response, however we suspect it was due to toxicity associated with membrane targeting and insertion of misfolded GFP. The Tat system is well known for its ability to tolerate proteins that have folded in the cytoplasm prior to export (DeLisa et al., 2003; Sanders et al., 2001). These folded proteins are inserted into the inner cytoplasmic membrane, often with assistance from molecular chaperones. When aberrantly folded proteins are targeted to the Tat system, cells exhibit severe growth defects that are likely due to proton leakage at the cytoplasmic membrane (Richter and Bruser, 2005). Hence, we suspect that the absence of the DnaK chaperone, which is reputed for its role in assisting the folding and translocation of Tat substrates (Oresnik et al., 2001; Perez-Rodriguez et al., 2007), caused ssTorA-GFP to misfold and become localized in a manner that was disruptive to membrane integrity.

The above explanation provides a conceptual link between our findings with a misfolded, mislocalized secreted protein and earlier studies where phage infection was the inducing cue for CRISPR-mediated silencing (Barrangou et al., 2007; Brouns et al., 2008). It is known that filamentous phage infection and overexpression of secreted or membrane proteins elicit overlapping stress responses (MacRitchie et al., 2008). Most notable among these is the phage shock protein (Psp) response that, like BaeSR, responds to membrane-damaging cues. For instance, the synthesis of an integral membrane protein encoded by phage gene IV, which is required at high levels during phage infection to facilitate virus assembly, is an inducing cue for the Psp signal transduction pathway (Brissette et al., 1990). Likewise, the Psp response has been shown to respond to defects in inner membrane protein translocation (Kleerebezem and Tommassen, 1993) and one of the key Psp response regulators, PspA, was shown to enhance protein export efficiency for

both Sec- (Kleerebezem and Tommassen, 1993) and Tat-targeted proteins including ssTorA-GFP (DeLisa et al., 2004). Based on these earlier findings, we speculate that activation of CRISPR-Cas silencing in response to protein secretion and phage infection is mediated by the same stress sensing mechanisms (Figure 17). While this minimally appears to involve BaeSR, we cannot rule out the involvement of additional envelope stress response pathways such as the Cpx and Psp systems that may work to cooperatively combat envelop stress by triggering CRISPR-Cas. We anticipate that comprehensive genetic analysis of the CRISPR-Cas system, including the underlying signaling mechanisms, should be possible by facile screening of cell fluorescence following ssTorA-GFP expression in *E. coli* cells.

Once activated, it appears that the CasA-E proteins form a complex called Cascade, and the CasE protein within the complex is responsible for processing full-length CRISPR RNA (Brouns et al., 2008). Here we showed instead that the formation of a ternary CasCDE core complex is necessary and sufficient for pre-CRISPR RNA processing. In our single-nucleotide resolution sequence gel analysis, evidence is quite strong that although CasC and CasE can cleave pre-CRISPR RNA individually, processing by the CasCDE core is significantly more specific and efficient (Table 2). The two discrete cleavage activities that we observed were assigned based on the behavior of the $CasE^{H20A}$ catalytic mutant. We conclude that CasE is responsible for making a specific cut to release the universal 5'-end in the mature crRNA, while CasC is responsible for generating the less homogeneous 3'-end of crRNA. The 5' and 3' boundaries of the resulting crRNA agree quite well with the *in vivo* cloning results (Brouns et al., 2008). A mechanistic model can therefore be generated to summarize our data, in which we hypothesize that by assuming a

defined tertiary structure, the CasCDE core complex possesses greater affinity and specificity for CRISPR repeat regions than individual Cas proteins. This in turn allows more efficient and highly specific pre-CRISPR processing by CasC and 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 also demonstrated the presence of a conserved CasCDE core in *T. thermophilus*. Our results are suggestive of a conserved pre-CRISPR processing complex. However, it should be noted that among the five Cascade proteins, only CasC and CasD are conserved across CRISPR-containing species. The other three proteins, CasA, CasB, and CasE, are not. Conversely, CRISPR repeat sequences and their predicted secondary structures can differ significantly between organisms (Grissa et al., 2008). Therefore, the pre-CRISPR processing in CasE-deficient organisms may differ significantly from that in *E. coli*, and the conserved CasCD heterodimer likely plays a pivotal role in pre-CRISPR processing in these organisms. 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 catalyze 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, and their pre-CRISPR processing mechanism remains unclear, underlining the diversity of CRISPR-Cas systems.

In contrast to the discrete RNase activities, we identified surprisingly abundant DNase activities among Cas proteins in *E. coli* K12 (Table 2). All

but Cas1 displayed a certain level of nuclease activity against either ds- or ss-DNA substrates. Unlike *Pseudomonas aeruginosa* Cas1 that was recently characterized as a metallo-nuclease with strong preference for Mn2+, but not Mg2+ as the cofactor (Wiedenheft et al., 2009), *E. coli* Cas1 is inactive under either Mg^{2+} or Mn^{2+} conditions (data not shown). Experiments in *E. coli* demonstrated that Cas1 and Cas2 likely function upstream of the CRISPR processing and DNA targeting stage, whereas Cas3 and an undefined number of Cascade proteins are required for efficient resistance against viral infection (Brouns et al., 2008). Our genetic screen showed that all five Cascade proteins are important for envelope stress-induced DNA silencing, yet biochemically only three proteins were required for CRISPR processing. It follows that CasA and CasB may participate in the downstream effector stage to target plasmid DNA for degradation (Figure 17). It remains unclear how crRNA guide the sequence-specific silencing of double-stranded phage genome DNA and, as in our case, circular plasmid DNA, but spares the CRISPR loci in chromosomal DNA. DNA methylation status is unlikely to be the cue because Cas protein activities are insensitive to methylation state of ds-DNA (data not shown). One possibility 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 or alternatively the chromosomal CRISPR loci may be coated by certain protein markers that prevent mis-targeting. There is evidence that mature crRNAs remain associated with the CasCDE complex (our unpublished observations), therefore upon hybrid formation between crRNA and target DNA, the same complex may initiate DNA degradation. Such a mechanistic model is reminiscent of the Argonaut protein in eukaryotic RNAi, in which Argonaut retains one strand of the small

interfering RNA (siRNA) as the guide to base-pair and cleave the target mRNAs. The absolute requirement of Cas3, a putative DEAD-box helicase fused to an HD nuclease is perplexing but not unencountered. A DExD/H-box helicase domain is also found in the Dicer protein and was shown to modulate its enzymatic activity in eukaryotic RNAi (Ma et al., 2008). Alternatively, instead of degrading the extrachromosomal DNA itself, the CasCDE complex may serve as a landmark to attract other effector proteins such as CasA and CasB through transient interactions (Brouns et al., 2008) (Figure 17).

Table 2. Summary of the DNase and RNase activity of Cas proteins and higher complexes.

	RNase	ds-DNase	ss-DNA
Cas1	-	-	-
Cas2	+++	-	+++
Cas3	-	-	++
CasA	-	+++	++
CasB	-	+	+/-
CasC	++	+	++
CasE	+++	+++	+++
CasCD	N/A	++	++
CasCE	+++	++	+++
CasCDE	+++	++	+++

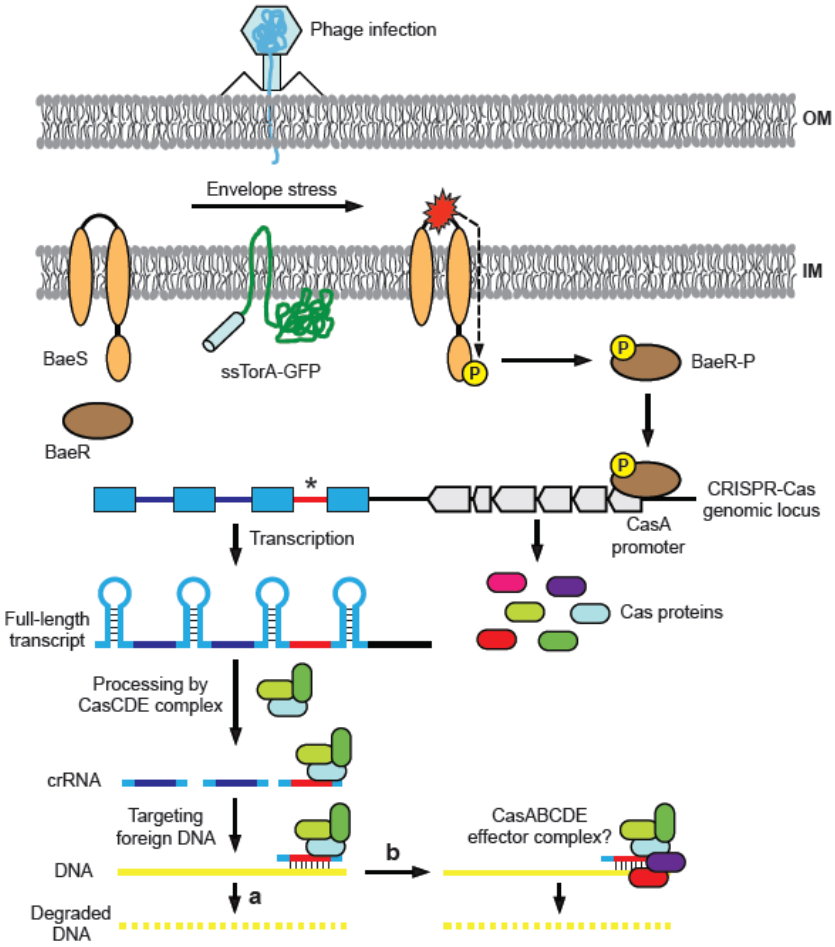


Figure 17. *Model of CRISPR-RNA mediated DNA silencing in E. coli.* Following an attack by a phage or the aberrant expression of a secreted or membrane protein, envelope stress response systems, such as the two-component BaeSR signal transduction pathway, are activated. An activated response regulator, 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 with its nicking and cleaving DNase activity (a). Whether CasCDE is a landmark to attract other effector proteins such as CasA and CasB through transient interactions (b) is currently unclear. 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.

4. Materials and Methods

4.1. Bacterial strains, plasmids and growth conditions

E. coli strain BW25113 (*lacI*^{q} *rrnB*$_{T14}$ D*lacZ*$_{WJ16}$ *hsdR514* D*araBAD*$_{AH33}$ D*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 *ΔdnaK* double knockout strains were created by P1*vir* phage transduction (see supplemental methods for details). Plasmid pTG encoded the *E. coli* ssTorA signal peptide fused to GFP in pTrc99A (Perez-Rodriguez et al., 2007) and pTrc-GFP encoded GFP cloned in pTrc99A (Diao et al., 2006). Additional strains and plasmids are described in the supplemental methods. Strains were routinely grown aerobically at 37°C in Luria-Bertani (LB) medium, and antibiotics were supplemented at the following concentrations: Amp, 100 mg/ml; Kan (5 mg/ml); spectinomycin (Spec, 50 mg/ml); tetracycline (Tet, 10 mg/ml). 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 hr prior to characterizing cellular fluorescence or harvesting plasmid DNA, total mRNA and total proteins. Cell viability and MIC were determined by non-selective plating and selective plating on Amp, respectively (see supplemental methods for details).

4.2. 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 (see supplemental methods for details). To isolate highly fluorescent clones,

library cells were sorted using a FACSCalibur flow cytometer (Becton Dickinson Biosciences) as described previously (DeLisa et al., 2002).

4.3 Cellular fluorescence analysis

Overnight cultures harboring GFP-based plasmids were subcultured into fresh LB medium with appropriate antibiotics and protein expression was induced in mid-exponential phase growth. After 4–6h, 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 analyzed 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).

4.4 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 described in the supplemental methods. 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 (BioRad), 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 1:2,500. 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.

4.5. Protein purification

Individual Cas proteins were expressed in His_6-tagged form and purified following standard protocols. CasCD, CasCE, and CasCDE complexes were co-expressed by co-transformation of plasmids encoding different Cas proteins and purified based on a His_6-tag on one of the proteins. All proteins and complexes were further purified on mono Q or Superdex 200 columns (see supplemental methods for details).

4.6. Electrophoretic mobility shift analysis (EMSA)

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 EMSA as described in the supplemental methods.

4.7. RNase activity assay

The RNA substrate was ^{32}P-labeled in a standard kinasing reaction and gel purified. RNA and proteins were incubated for 30 min in a buffer containing 20 mM Tris pH 7.5, 40 mM NaCl, and 2.5 mM $MgCl_2$. The reaction was quenched by phenol extraction, and the RNA was EtOH precipitated and separated on a 16% sequencing gel. After overnight exposure, the image plate was scanned with a Typhon 9200.

4.8. 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 *Rsa*I (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.

4.9. DNase activity assay

The ds-DNase assay was carried out in a buffer containing 50 mM Tris-HCl pH 7.9, 100mM NaCl, 10mM $MgCl_2$, 1mM dithiothreitol, 1 mg/ml BSA. Supercoiled pUC19 plasmid was purified using a Qiagen Megaprep kit. This plasmid was digested with *Bam*HI restriction enzyme to produce linearized DNA substrate. The reaction was then phenol-extracted and separated on a 0.8% agarose gel. The ss-DNase assay was carried out in the same buffer with the addition of the same amount of proteins. A 5' HEX-labled 29-nt DNA oligo encoding part of spacer 5 was used as substrate to allow fluorescent detection. After incubating 10 nM substrate with 1 μM Cas proteins for 30 min in a 20-μl reaction, the reactions were separated on a 10% urea-TBE acrylamide gel and the fluorescent signals were scanned with a Typhon 9200 Imager.

5. Supplemental Methods

5.1. Bacterial strains and plasmids

BW25113 D*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 *ΔdnaK* cells that had the Kan resistance marker in *dnaK* eliminated as described (Datsenko and Wanner, 2000). BL21(DE3) (F^- *ompT hsdS*$_B$ (r_B-m_B-) *gal dcm* l(DE3)) (Novagen) or T7 Express *lysY* (New England Biolabs) cells were used as hosts for expression of the BaeR and Cas enzymes, respectively.

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 (Castanie 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 pssS8-GFP was created by mutating the C57 and C60 nucleotides in *sstorA* to T57 and T60, respectively. Plasmid pss5.8.12-GFP was created by further mutating pssS8-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. Individual *cas* genes were amplified from *E. coli* K12 genomic DNA. For individual protein expression, CasC, CasE and $CasE^{H20A}$ were 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 co-expression 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. BaeS and BaeR were cloned in pET-21a(+) (Novagen) with a C-terminal His_6-tag.

5.2. Generation of transposon library and mapping of insertions

The initial transposon insertion library in NLL51 was constructed as described previously (Takiff et al., 1989) and contained $4.2x10^4$ transductants, representing ~10-fold coverage of the genome. The mini-Tn*10* insertion library was then transduced in *ΔdnaK* cells carrying pTG by P1*vir* phage transduction and $3.0x10^4$ clones were recovered on LB-Tet/Amp plates, for ~8-fold coverage of the genome. To isolate highly fluorescent clones, library cells were grown overnight, subcultured and induced for 4-6 hr. 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 et al., 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.

5.3. 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 mL (CFU/mL) 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 was determined by counting individual colonies on plates after overnight growth. Reported CFU/mL values were the average of three replicate experiments.

5.4. Northern blot analysis

Northern blots were performed as follows: 5 mg 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 (DIG)-labeled 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 fragmented salmon sperm DNA and 5 µg/ml yeast tRNA). Membranes were washed and exposed to anti-DIG 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.

5.5. 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 hr 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, 5mM 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 5 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 (GE Healthcare). Cas2 was expressed and purified in an N-terminal GST-

fusion form. All of the purified proteins were analyzed by SDS-PAGE and judged to be greater than 95% purity.

To co-express the CasCE complex, pCDFDuet-1-CasC and pET-21-CasE were co-transformed into T7 Express *lysY* cells. Following co-expression, the CasCE complex was purified on a Ni-sepharose column followed by a Superdex 200 sizing column on FPLC as described above. The fact that CasC co-purified 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 co-expressed 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 co-transformed 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 co-expression, the CasCDE complex was purified with a Ni-sepharose column followed by a Superdex 200 column on FPLC as described above. Although only CasD contains a His_6-tag, CasE and CasC co-purifies with CasD under stringent purification conditions, suggesting that the 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 co-expression 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 co-purified on the Ni-sepharose column

under stringent purification conditions, and CasCDE and CasCD were separated on Superdex 200 column based on their size difference.

5.6. Electrophoretic mobility shift analysis (EMSA)

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 hr at 30°C. Next, 0.15 pmol of DNA was mixed with varying amounts (9.2 μM, 11.5 μM, 16.1 μM, 24.2 μM, 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 (BioRad) 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 (BioRad).

6. Acknowledgements

We thank Joseph Peters and Adam Parks for assistance with transposon library construction and mapping of insertion mutations. We also thank Changrui Lu and Matthew Bratkowski for helpful discussions. R.P.-R. gratefully acknowledges Cornell University for support through a Sloan Fellowship. This material is based upon work supported by the National Science Foundation under grant no. 0449080 (to M.P.D.), a NYSTAR James D. Watson Award (to M.P.D.), a NYSTAR Distinguished Faculty Award (to M.P.D.).

CHAPTER 4

PRELIMINARY EXPERIMENTS AND FUTURE DIRECTIONS

1. Reconstitution of gene silencing of ssTorA-gfp in BL21(DE3) cells through addition of CRISPR-Cas system.

In the previous chapters, the experiments depicted were performed in *E. coli* K12 cells, which have a CRISPR-spacer locus adjacent to the endogenous *cas* genes. In our next set of experiments, we wanted to reconstitute the silencing of ssTorA-GFP through the addition of the CRISPR-Cas system in a *E. coli* strain that lacks this machinery. To address this, we expressed the five subunits of the Cascade complex (*casABCDE*), *cas1*, and *cas2* in *E. coli* BL21(DE3) cells which lack endogenous *cas* genes and the CRISPR-spacer locus (Brouns et al., 2008). As preliminary experiments, the *cas* genes, were overexpressed as: (1) a whole operon, (2) individual subunits, and (3) the core effector complex CasCDE, each together with *ssTorA-gfp*.

First, we started by expressing ssTorA-GFP and the entire *Cas* operon from a plasmid (p*casABCDE12*, provided by John van der Oost (Brouns et al., 2008)) in BL21(DE3) and BL21(DE3) *ΔdnaK* cells. To our surprise, the plasmid carrying ssTorA-GFP was almost completely silenced in both wild-type and *ΔdnaK* mutants (see Figure 18). Likewise, a control plasmid carrying only GFP without the TorA signal sequence was similarly targeted when was expressed along with the Cas operon plasmid in the same cell lines (see Figure 19). BL21(DE3) cells lack the CRISPR-Spacer DNA locus, and without the guide RNAs the overexpression of all Cas components results in non-

specific degradation of plasmid DNA either in the presence or absence of the chaperone DnaK. This is entirely consistent with our *in vitro* experiments where we showed that all Cas proteins (except for Cas1) had endodeoxyribonuclease activities but they were not site specific, as using pssTorA-GFP or an unrelated pUC19 plasmid did not alter the experimental outcome (see Figure 16). Thus, we can suggest that the guide CRISPR-RNAs are critical to ensuring target specificity.

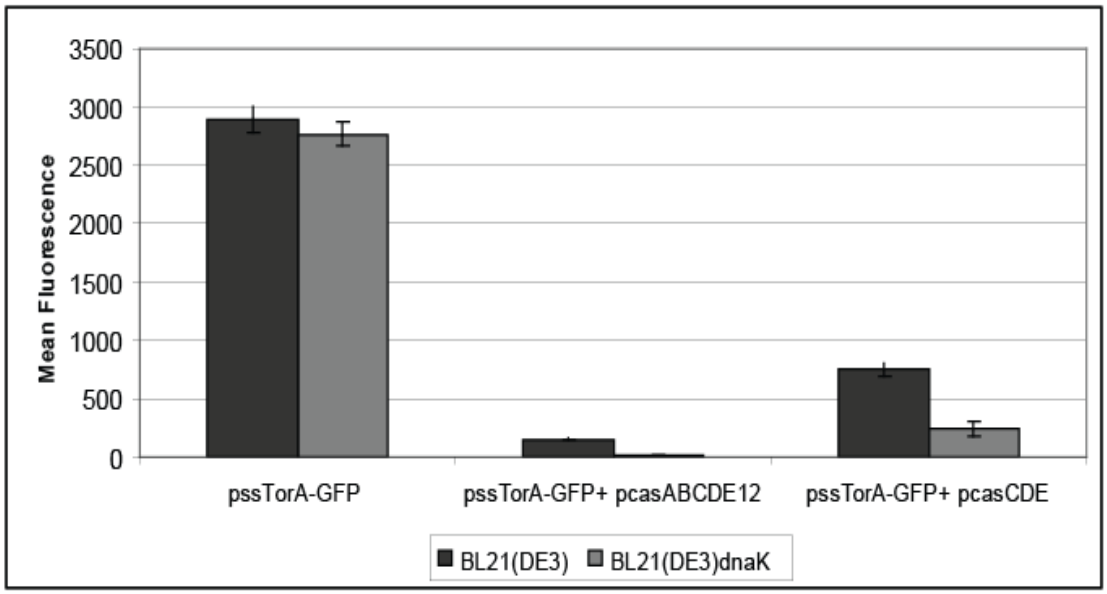


Figure 18. *Overexpression of the Cas operon and the Cas core complex leads to a non-specific DNA-silencing in wild-type and ΔdnaK mutants expressing ssTorA-GFP*. Flow cytometric screening of *E. coli* BL21(DE3) (wt) and BL21(DE3) *ΔdnaK* expressing ssTorA-GFP, CasABCDE12, and CasCDE, as indicated.

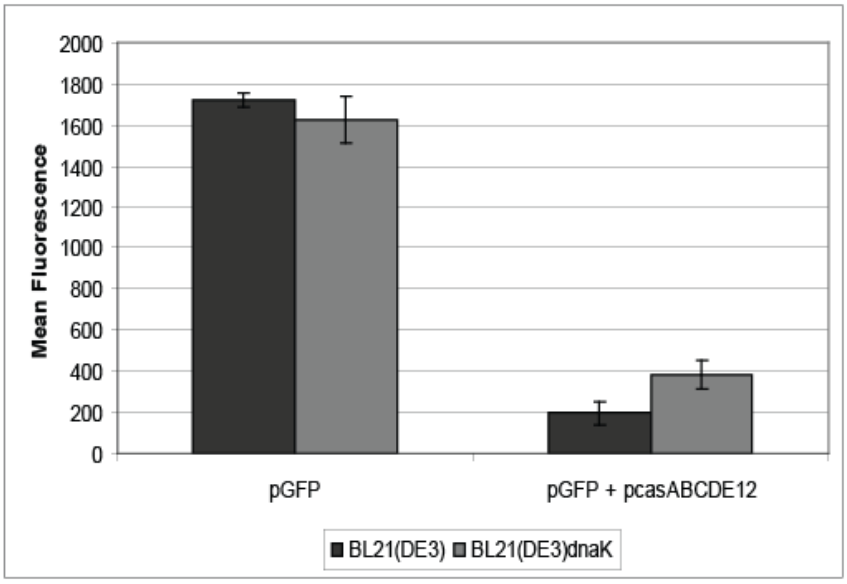


Figure 19. *Overexpression of the Cas operon and the Cas core complex leads to a non-specific DNA-silencing in wild-type and ΔdnaK mutants expressing GFP*. Flow cytometric screening of *E. coli* BL21(DE3) (wt) and BL21(DE3) *ΔdnaK* expressing GFP and CasABCDE12, as indicated.

To see if the same non-specific DNA degradation can be accomplished by individual Cas proteins, we next overexpressed each one of these components individually in wild-type and *ΔdnaK* cells together with ssTorA-GFP. In general, the overexpression of independent Cas proteins in wild-type and *ΔdnaK* mutants did not affect the production of ssTorA-GFP, suggesting that the protein products of the cascade *casABCDE* and *cas1* may act together for non-specific DNA targeting and degradation (see Figure 20). This is in contrast to our *in vitro* experiments where we observed DNAse activity from each of the Cas components, except for Cas1 (see Chapter 3). Interestingly, the expression of ssTorA-GFP fluorescence was almost null when Cas2 was expressed in *ΔdnaK* mutants, implying a possible individual

DNA cleavage activity by this enzyme when the chaperone DnaK is not present.

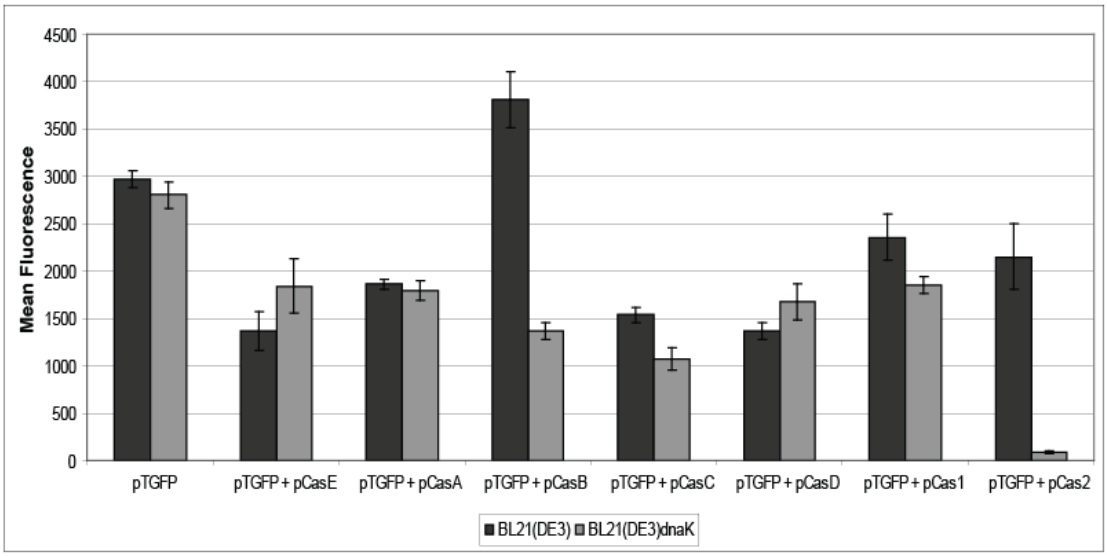


Figure 20. *Overexpression of individual Cas proteins in wild type and ΔdnaK mutants does not affect expression of ssTorA-GFP*. Flow cytometric screening of *E. coli* BL21(DE3) (wt) and BL21(DE3) *ΔdnaK* expressing ssTorA-GFP (pTGFP), CasE, CasA, CasB, CasC, CasD, Cas1, and Cas2, as indicated.

These results suggest that almost all Cas proteins, when highly expressed in both wild type and *ΔdnaK* mutants of cells that do not contain natural CRISPR-Cas machinery, act together and not individually in the non-specific targeting of the plasmid DNA. However, we have shown *in vitro* that CasC, D, and E may form a ternary protein complex and is sufficient for DNA degradation in *E. coli* K12 (see Chapter 3). To test if this is also true for BL21(DE3) cells *in vivo*, we overexpressed these three proteins (p*casCDE*, provided by John van der Oost (Brouns et al., 2008)) in wild-type and *ΔdnaK* mutants of BL21(DE3). When the Cas C, D, and E proteins are expressed at high levels in both strains, the fluorescence of GFP decreases significantly, showing a similar non-specific DNA degradation as was observed when all

Cas proteins were overexpressed in BL21(DE3) cells (see Figure 18). Moreover, this result implies that high levels of proteins Cas C, D, and E, are enough for non-specific degradation of foreign DNA even though the guide CRISPR-RNAs are not present in these cells. Similar to *E.coli* K12, overexpressed Cas proteins other than C, D, and E in BL21(DE3) cells might produce the non-specific DNA targeting through transient interactions with the CasCDE complex at the final stage of degradation (see DNA-silencing mechanism proposed in Chapter 3).

2. Future Directions

2.1. Reconstitution of DNA silencing in BL21 (DE3) cells by low expression of Cas proteins and CRISPR-spacers RNAs.

We stated above the hypothesis that the overexpression of the all *cas* genes together, or specifically *casCDE*, leads to non-specific DNA-targeting in the absence of CRISPR-RNAs. To test if high levels of Cas proteins are the cause, the *cas* genes can be cloned into a promoterless low copy plasmid (e.g., pFZY1, one to two plasmids per cell) (Koop et al., 1987). The *cas* genes would be transcribed with their individual natural promoters at their endogenous levels. Similarly to the experiments described above, the *cas* genes can be cloned as the entire operon (*casABCDE, cas1, and cas2*) or as the Cas core complex (*casCDE*) in pFZY1. These two plasmids can be individually transformed into BL21(DE3) and in BL21(DE3) *ΔdnaK* mutants together with ssTorA-GFP. If DNA-silencing of GFP is not observed when these proteins are expressed at very low levels, we could say that our hypothesis is correct. We can also suggest that Cas proteins, when

expressed at very low levels, could potentially need the guide CRISPR-RNAs to target plasmid DNA by specificity in the absence of the chaperone DnaK. To test this assumption, a plasmid encoding the CRISPR-spacers DNA can also be cloned and transformed in BL21(DE3) cells. In Chapter 3, we discussed that Spacers 5, 8, and 12, provide high affinity to ssTorA and lead to the degradation of GFP in *E. coli* K12. Therefore, another promoterless low copy plasmid (e.g., pCS26, (Harmon and McKay, 1987)) encoding these three spacers in the order of: CRISPR-Spacer5-CRISPR-Spacer8-CRISPR-Spacer12-CRISPR could be designed. This particular arrangement is based on our proposed structure of a mature crRNA (a single CRISPR-Spacer-CRISPR unit), where the 5'- and 3'- cleavage sites resides only in the repeats and not in the spacers regions (Chapter 3). This plasmid would be transformed into wild type and *ΔdnaK* mutants, together with ssTorA-GFP and pFZY1 carrying either the *cas* operon or the Cas core complex. If our assumption is correct, expression of GFP will decrease significantly only in *ΔdnaK* mutants and not in wild type cells, indicating that we reconstituted the CRISPR-RNA mediated DNA silencing in BL21(DE3) cells.

If these genetic experiments performed in BL21(DE3) cells produce the expected results, we would be able to conclude that (1) the overexpression of Cas proteins leads to non-specific DNA targeting and degradation without the need of the CRISPR-RNAs guides, either in the absence or presence of the molecular chaperone DnaK, (2) there are naturally a few "free" Cas components in the cells, otherwise it would be quite harmful to plasmid maintenance, and (3) when expressed at low levels, the Cas proteins act together with CRISPR-RNAs to degrade plasmid DNA and these guide RNAs

are critical to ensure target specificity, just as we have proposed in our studies in *E. coli* K12.

2.2. Additional genetic validation and biochemical studies in E. coli K12.

2.2.1. Dependence of phenotype on chromosomal CRISPR-Spacer DNA.

Our next priority is to study the effect of the complete deletion of the entire CRISPR-Spacer locus or a single CRISPR-spacer unit from the *E. coli* genome. The CRISPR-Spacer locus is an important part of the DNA silencing system, since it is believed that the CRISPRs, together with the spacers provide the specificity for DNA targeting and thus, developing an immunity system in cells (Barrangou et al., 2007). To perform this deletion, the phage λ lambda red recombination method (Datsenko and Wanner, 2000) can be employed (see Figure 21). In this case, the chloramphenicol (Cm)-coding *cat* gene could be chosen to select for the deletion to avoid redundancy of antibiotic resistance.

These deletions should be made in wild type *E. coli* K12 and in *ΔdnaK* mutants. The absence of the CRISPR-Spacer DNA region is likely to disturb the DNA-silencing found in *ΔdnaK* mutants and we would expect that the level of fluorescence of ssTorA-GFP will be the same as wild-type cells. This result would confirm our hypothesis that even though the endogenous Cas proteins are expressed in these mutants, these proteins need the CRISPR-RNAs to trigger silencing. Also, we should pursue the deletion of specific CRISPR-Spacers units 5, 8, and 12 that we found have a high binding affinity to the ssTorA sequence. Following the same procedure of deletion and fluorescence analysis, we also expect to observe an increase in fluorescence in *ΔdnaK*

mutants. This would mean that the absence of any of these spacers can decrease binding to ssTorA, thus, weakening the trigger for DNA-silencing.

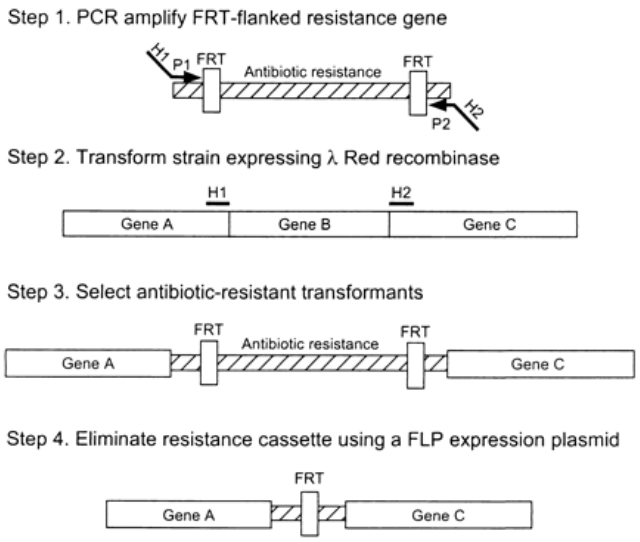


Figure 21. *A simple gene disruption strategy***.** H1 and H2 refer to homology extensions or regions, and P1 and P2 refer to priming sites (Datsenko and Wanner, 2000).

2.2.2. Dependence of phenotype on ssTorA.

The next focus of this research should be based on the homology studies done with spacer sequences and *ssTorA* (see Chapter 3). These results revealed strong binding interactions between three candidate spacer regions and ssTorA, and led us to hypothesize that this DNA-silencing mechanism specifically targets ssTorA. Since all the experiments performed to test the targeting of ssTorA containing protein used GFP as the reporter, it would be interesting to clone ssTorA in front of other substrates, such as other native (e.g., CueO and SufI) or non-native Tat-targeted proteins (e.g., ssTorA-BlaC) in *E. coli* K12. These Tat-targeted proteins can be transformed individually in wild type and *ΔdnaK* cells, and tested for their production by

immunoblotting subcellular fractions or by measuring cell growth (e.g. ssTorA-Bla). The rationale for doing these experiments is to be able (1) to relate the DNA-silencing with the targeting of spacers to ssTorA fused to other proteins and (2) to show that the DNA-silencing phenomenon is not GFP specific. The decrease in protein production in *ΔdnaK* mutants would suggest that, in addition to ssTorA-GFP, DNA-silencing becomes active due to membrane insertion of misfolded Tat-targeted proteins in the absence of DnaK chaperone activity, and when genes are downstream of *ssTorA*.

2.2.3. Cas protein subunits interactions in vivo

In our studies, we have found that CasC interacts with CasE and CasD individually or with both at the same time. However, these studies were made *in vitro* (see Chapter 3) and we find it important to study their native interactions inside *E. coli* cells. Our first priority for an *in vivo* experiment is to adapt the protein fragment complementation assay (PCA), which is based on engineered split reporter protein fragments that do not interact or have functional activity alone (Nyfeler et al., 2005). Both CasC and CasD proteins can be tagged with two split yellow fluorescence protein (YFP) fragments and cloned into pBAD vectors (see Figure 22). If CasC and CasD interact specifically *in vivo*, they will bring the two split YFP fragments together where they will fold into a fully functional protein and give a fluorescence signal, which could be analyzed using a flow cytometer. In this manner, cells could be transformed with plasmids carrying YFP1-CasC and CasD-YFP2, and a positive control could the high-affinity Fos/Jun leucine zippers fused to the split YFP fragments, and the negative control could be the non-interacting split YFP fragments alone. The same procedure can be done for CasC and CasE as

well. If these experiments show that these Cas proteins interact *in vivo*, we will have even stronger evidence that they interact while binding to CRISPR-RNAs to target foreign DNA.

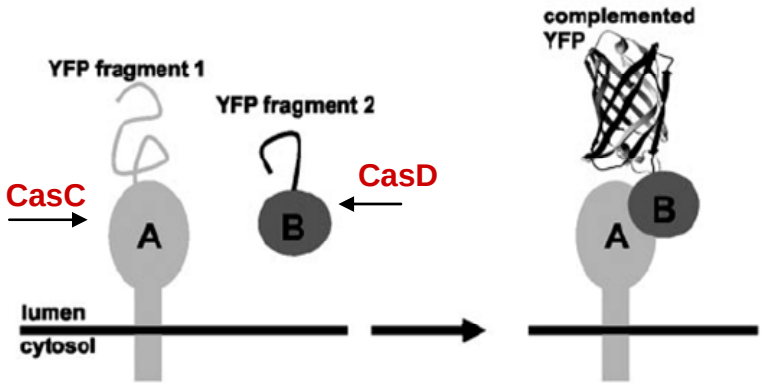


Figure 22. *YFP fragment complementation in the lumen of the secretory pathway.* YFP fragments 1 (amino acids 1-158) and 2 (amino acids 159-239) are fused to a protein A to a soluble protein B. Homooligomerization of the membrane protein A or its interaction with B brings the two fragments of YFP into close proximity and leads to complementation into functional fluorescent YFP by folding into an active 3D structure (Nyfeler et al., 2005).

Another approach that could be used for studying Cas proteins interactions *in vivo* is to perform formaldehyde crosslinking experiments using lysates generated from *E. coli* K12 wild type and *ΔdnaK* cells individually expressing a chosen pair of proteins. This would be likely to work, since using this method, we have previously shown strong individual interactions between the chaperones DnaK and GroEL with the native Tat substrate TorA (Perez-Rodriguez et al., 2007).

2.3. Another direction

As stated in the previous chapter, in the search for factors that could lead to degradation of ssTorA-GFP in *ΔdnaK* mutants we screened a mini-Tn10 library to isolate suppressor mutants that restored expression of GFP. As a result, we found 8 transposon hits and they are summarized in Table 3.

Table 3. Isolated mutants that restored expression of GFP in *ΔdnaK* mutants.

Gene name	Function
xapA	purine nucleoside phosphorylase II
glgX	glycogen operon
cysH	Enzyme –sulfur metabolism
yghQ	Predicted inner membrane, serine protease
ybfH	Predicted protein
ygjC	Predicted tartrate: antiporter
ybhI	Putative transporter: function unknown
ygcH or casE	CRISPR-associated protein

As we expected, we found that one of the suppressors mapped to the *yghQ* gene, which is a predicted inner membrane serine protease. This protease could be responsible for the degradation of misfolded proteins in *ΔdnaK* mutants, whereby the absence of this chaperone promotes its proteolytic activity. We did not decide to follow up on this, since we thought that *casE* was more interesting for its potential role in a novel prokaryotic RNAi system (Makarova et al., 2006). Little is known about *yghQ*, and it would be interesting to investigate the possible role of this protease in the housekeeping of misfolded proteins in the absence of DnaK.

CHAPTER 5

SUMMARY, CONCLUSIONS AND ENGINEERING APPLICATIONS

1. Summary and Conclusions

Originally, we sought to determine and identify factors that, in addition to the TatABC machinery, would enhance Tat transport. In light of some evidence that general chaperones (e.g., DnaK and GroEL) and proteases (e.g., ClpXP, FtsH) play a role in the early stages of Tat translocation in bacteria (Oresnik et al., 2001; Rodrigue et al., 1996; Xu et al., 2005) and plants (Dionisi et al., 1998; Leheny et al., 1998; Madueno et al., 1993; Molik et al., 2001), we systematically tested a collection of *E. coli* chaperone mutants for their affect on transport of the Green Fluorescent Protein (GFP). Specifically, we found that the molecular chaperone DnaK exerted the most significant effect on the abundance and transport of synthetic (e.g., ssTorA-GFP, ssTorA-GFP-SsrA, and ssTorA-MBP) and natural Tat substrates (e.g., CueO and SufI). Moreover, we found that the effect of this chaperone is not entirely dependent on its co-chaperones DnaJ and GrpE, since their absence only resulted in a minor drop in Tat substrate stability. Even though DnaK was only previously linked to cofactor-containing Tat substrates, we observed that even Tat substrates lacking cofactors showed instability in the absence of this chaperone. This suggests that the role of DnaK in maintaining the stability of Tat substrates is independent of cofactor biogenesis.

We envision that DnaK interacts with Tat substrates in a highly coordinated fashion and functions to stabilize such substrates against misfolding and proteolysis. Furthermore, it may serve as an escort for guiding Tat substrates to the TatABC translocon while preventing early engagement of

the Tat machinery. The observation that overexpressed DnaK improved the expression and translocation efficiency of Tat-targeted substrates reinforced the connection between DnaK and the Tat transport process. We speculate that this improvement is directly related to: (1) the protection of Tat substrates from cytoplasmic degradation or early targeting to the translocon, and (2) the increase of the overall fraction of correctly folded proteins in the cytoplasm. These results are significant from a biotechnology standpoint, since they provide a strategy for enhancing the Tat system as a platform for protein expression and secretion.

Following these observations, we decided to study the cause of rapid protein degradation in *ΔdnaK* mutants in more detail. Our initial hypothesis was that ssTorA-GFP was unstable in the absence of DnaK and, thus, more susceptible to proteolytic degradation. However, when we analyzed the GFP mRNA levels by Northern blot, *⊿dnaK* mutants revealed very low levels of mRNA relative to wild type. In an effort to find the main component(s) responsible for silencing ssTorA-GFP mRNA, we constructed and screened a mini-Tn10 transposon library in a *⊿dnaK* strain background. The purpose of making this library was to isolate mutants that restore expression and, thus, fluorescence of ssTorA-GFP. Strikingly, one of the mini-Tn10 gene knockouts that we isolated mapped to a gene named *casE* (formerly named *ygcH*). At the time, *casE* was thought to encode a hypothetical protein with unknown function in the genome of *E. coli* K-12. However, this protein is now proposed to be an endoribonuclease in the *cas* genes family and part of a possible prokaryotic RNAi system (Brouns et al., 2008).

After studying the existent models of a putative RNAi system in bacteria, we found that *casE* is part of the Cascade system, *casABCDE* in the

genome of *E. coli* K-12 (Brouns et al., 2008). This system, together with their adjacent DNA, CRISPR-spacers loci, protects bacteria against phage infection (Barrangou et al., 2007; Brouns et al., 2008) and plasmid conjugation (Marraffini and Sontheimer, 2008) by targeting extrachromosomal DNA by sequence specificity. Yet, we unexpectedly uncovered evidence that the CRISPR-Cas system can also protect bacteria from protein mislocalization. This seemed especially true when the integrity of the inner (cytoplasmic) membrane is threatened. We now hypothesize that ssTorA-GFP accumulates in the inner membrane of *ΔdnaK* cells and the resulting toxicity may induce an envelope stress response system and will activate the Cas proteins to silence the plasmid DNA. We also showed that such silencing is dependent upon a BaeSR signal transduction pathway, indicating for the first time, that an intracellular mechanism is also involved in activating the CRISPR-Cas system. Although we are still uncertain if or why ssTorA-GFP caused such a stress response in *ΔdnaK* mutants, we suspect that the absence of this chaperone, known to assist in folding and translocation of Tat substrates (Oresnik et al., 2001; Perez-Rodriguez et al., 2007), might cause the misfolding and the detrimental membrane accumulation of ssTorA-GFP.

Based on the results of this study, we propose a model for the CRISPR-RNA mediated DNA silencing in *E. coli* K-12 (see Figure 17). Followed by an aberrant expression of a secreted or membrane protein, an envelope stress response system, such as the two component transduction pathway BaeSR is activated. Phosphorylation of BaeR by BaeS triggers expression of the Cas proteins and perhaps the transcription of the CRISPR-spacer region. Since we showed that the CasCDE complex is specific and efficient in processing pre-crRNAs, we hypothesized that this protein complex binds to the CRISPR

region by assuming a defined quaternary structure. In addition, we showed that all five Cascade proteins have DNase activity and are all necessary for envelope stress-induced plasmid silencing. It remains obscure how the crRNAs target the foreign double stranded DNA of phages and plasmids and do not target 'self' DNA. However, we suspect that transcription of phage or plasmid DNA (replication fork) serves as an opportunity for crRNAs to bind to the single-stranded DNA by base pairing. Another possibility could be that the CRISPR locus becomes armored in the chromosome by protein markers that prevent mis-targeting of crRNAs.

Lastly, we have proposed in this model, that CasC, CasD, and CasE attract other effector proteins (e.g. CasA and CasB) through transient interactions. With the help of the effector Cas complex (CasABCDE), crRNAs will pair to the DNA target during plasmid or phage DNA replication. Interestingly, some bacteria acquire new spacers in their chromosome which are derived from the invading DNA, resulting in the resistance and survival by CRISPR-RNA mediated degradation of foreign DNA (Brouns et al., 2008).

In this study, we provided the mechanistic insights: (1) Cas proteins are triggered by the envelope response system BaeSR, and (2) pre-crRNAs are processed into mature crRNAs that will serve as guide sRNAs for targerting DNA. Nonetheless, questions remain as to how exactly crRNAs target extrachromosomal DNA and how bacteria acquire immunity through integration of short pieces of such DNA.

2. Engineering Applications

2.1. RNA interference and antisense technology

2.1.1. Eukaryotic cells

Besides unraveling the origins and detailed mechanism of RNA interference and using RNAi to screen for genomic functional analysis (Echeverri and Perrimon, 2006), the creation of new delivery methods for genetic materials into cells and animals has become very important (Hebert et al., 2008). RNAi is arising to be a new therapeutic tool for combating a wide range of diseases, including cancer, viral infection, and neurodegenerative diseases (de Fougerolles et al., 2007; Dykxhoorn et al., 2006). Though great potential exists, there are many complexities involved including the effective delivery of the RNAi molecules into the target cell and the quantification of possible side effects (Hebert et al., 2008).

One of the applications of the RNAi technology could be in metabolic engineering to alter genes and regulatory structures in order to achieve a desired phenotype (Bailey, 1991; Stephanopoulos and Vallino, 1991). In addition to making the molecule of interest, the goal of metabolic engineering can include desired phenotypes such as survival in adverse conditions (e.g., when experiencing a decrease in oxygen and nutrients) (Rao and Bredesen, 2004). Typically, this involves alteration of central metabolism so flux through specific pathways is increased while others are decreased (Hebert et al., 2008). Often this is accomplished through the overexpression or knockout of specific genes, but instead of this approach, alteration of the central metabolism could also be accomplished by transient silencing of desired genes via RNAi.

An emerging interest in metabolic engineering is apoptosis, or programmed cell death. Up to 80% of cell death has been observed in bioreactors leading to a decrease in yield and quality of the product (Majors et al., 2007). To address this issue, an RNAi strategy has been implemented whereby proteases in CHO cells are targeted and downregulated. This silencing mechanism has resulted in a higher yield of recombinant proteins (Lai et al., 2004; Wong et al., 2006).

Another example of metabolic engineering where RNAi could be applied is in the cycle of cell growth. After growing rapidly in early stages, the cells go through a phase called G1 where their growth is very slow but their productivity in very high (Gammell et al., 2007). Although several eukaryotic cell lines have undergone metabolic engineering to control the cell growth cycle, RNAi has been shown to increase cell growth and protein production simultaneously. For example, a study made in *Drosophila* S2 cells revealed that dsRNA against the cell growth controller *cycE* (a gene encoding the Cyclin E protein that mediates the progression from G1 to the DNA replication phase (S) of the cell cycle), increased cell growth and, at the same time, enhanced the production of desired recombinant proteins (March and Bentley, 2006; March and Bentley, 2007).

2.1.2. Prokaryotic cells

Aside from eukaryotic cells, the RNAi mechanism has been shown to work well in bacterial systems as a metabolic engineering tool and a potential therapeutic agent (Hebert et al., 2008). Antisense molecules, which anneal to complementary mRNAs to block translation or induce rapid degradation of proteins, could be used to increase productivity of desired proteins in bacterial

hosts. Also, as discussed previously, (see Chapter 3), these antisense approaches have been proposed to protect bacteria against foreign DNA.

Antisense molecules can also be used for reducing pathogenicity in bacteria, either by weakening the potency of toxins or virulence mechanisms, or by targeting antibiotic resistance genes. Studies have shown that small antisense sRNAs expressed in plasmids can be made to pair with the target mRNA (e.g., encoding toxins and antibiotic resistance genes) and activate cleavage by RNAse P activity (Guerrier-Takada et al., 1995; Soler Bistue et al., 2007). Though this concept is very promising from a therapeutic perspective, it is still limited due to the fact that antisense RNAs are more effective when expressed from plasmids (Hebert et al., 2008). Clearly, targeted delivery of plasmids to bacteria in the human body would limit this strategy in the clinic. In response to this problem, researchers have chemically synthesized short DNA sequences harboring antisense phosphodiester oligodeoxyribonucleotides (asODNs) (Hebert et al., 2008). These oligoribonucelotides are designed to be complementary to the mRNA of interest with the goal of reducing antibiotic resistance in cells by inducing the RNAse P digestion of such mRNA. However, due to the fast degradation of asODNs by nucleases, synthetic analogs such as phosphorothiate ODNs (PS-ODNs), peptide nucleic acids (PNAs), and phosphorodiamidate morpholino oligomers (PMOs) have also been made (Hebert et al., 2008). The best studied analog is the PS-ODN which has been shown to inhibit growth in *Mycobacterium tuberculosis* (Harth et al., 2007), and to restore sensitivity in an antibiotic resistance strain *Staphylococcus aureus* (Meng et al., 2006).

In general, the concept of using antisense molecules to target mRNAs could have a great potential in antimicrobial therapies, but not until a

mechanism for an effective and specific cellular delivery is completely developed. Furthermore, new antisense technologies will have to be continually adapted, as bacteria will continue to evolve new mechanisms of resistance to therapeutic methods.

2.2. CRISPR-RNAs technologies

CRISPRs, together with spacers and Cas proteins, may confer protection to the bacterial cells by targeting and degrading invading DNA possibly by an RNAi-like mechanism (our own observations). But, interestingly, more than a decade before the discovery that these repeats participate in attacking phage DNA, they were used for strain genotyping (Groenen et al., 1993). Specifically, CRISPRs have been very useful for strain detection of *M. tuberculosis*, which is known to constantly evolve by changing the number of repeats and spacers in its genome (Kamerbeek et al., 1997). Using CRISPRs as priming templates, the exact DNA sequence of the spacers is obtained, and thus, used to characterize and catalogue clinical strains of *M. tuberculosis*.

Another interesting application of CRISPRs can be found in the diary industries. Since most of the dairy industry uses bacteria to produce various foods (e.g., yogurt, cheese, etc.), there is often a need to confront phage infection. Phage infection results in the failure of many culture batches and can lead to high economic losses. CRISPRs can be used to defeat this problem. For instance, spacers derived from the DNA of some of the phages that commonly attack the cultures, can be artificially added to the genome of bacteria (Sorek et al., 2008). Thus, these engineered cells have been conferred immunity to survive infection when faced with a new attack from

known phages. This application has already been patented by Horvarth and co-workers (International patent # 2007025097, 2007.).

As observed in our work (see Chapter 3), CRISPR-RNAs bind to target DNA by sequence specificity. Based on this, it has been proposed that genes could be silenced by expressing a plasmid carrying part of the target chromosomal gene DNA as spacers instead of knocking out genes directly in the bacterial chromosome. The design of this CRISPR-spacer locus in this plasmid would have to be comprised of a complete sequence in the order CRISPR, spacer, and then CRISPR. The cloning of this locus has been somewhat difficult by traditional molecular biology techniques since the repetitive nature of the CRISPRs leads to many non-specific cloning products (our own observations).

It is envisioned that CRISPRs will play an important role in microbial genetics in the future, since various genes can also be knocked down simultaneously due to the repeats encoded by CRISPRs (Sorek et al., 2008). However, the exact mechanism of inserting new spacers in the bacterial chromosome and the mechanism of targeting invading DNA, are questions that need to be answered before considering the CRISPR-RNAs technology for further applications.

REFERENCES

Aiba, H. (2007). Mechanism of RNA silencing by Hfq-binding small RNAs. Curr Opin Microbiol *10*, 134-139.

Alami, M., Trescher, D., Wu, L. F., and Muller, M. (2002). Separate analysis of twin-arginine translocation (Tat)-specific membrane binding and translocation in Escherichia coli. J Biol Chem *277*, 20499-20503.

Alba, B. M., and Gross, C. A. (2004). Regulation of the Escherichia coli sigma-dependent envelope stress response. Mol Microbiol *52*, 613–619.

Angelini, S., Moreno, R., Gouffi, K., Santini, C., Yamagishi, A., Berenguer, J., and Wu, L. (2001). Export of Thermus thermophilus alkaline phosphatase via the twin-arginine translocation pathway in Escherichia coli. FEBS Lett *506*, 103-107.

Arsene, F., Tomoyasu, T., and Bukau, B. (2000). The heat shock response of *Escherichia coli*. International Journal of Food Microbiology *55*, 3-9.

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.

Babitzke, P., and Romeo, T. (2007). CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr Opin Microbiol *10*, 156-163.

Bailey, J. E. (1991). Toward a science of metabolic engineering. Science *252*, 1668-1675.

Baranova, N., and Nikaido, H. (2002). The baeSR two-component regulatory system activates transcription of the yegMNOB (mdtABCD) transporter gene cluster in Escherichia coli 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.

Barrett, C. M., Ray, N., Thomas, J. D., Robinson, C., and Bolhuis, A. (2003). Quantitative export of a reporter protein, GFP, by the twin-arginine translocation pathway in Escherichia coli. Biochem Biophys Res Commun *304*, 279-284.

Becker, L. A., Bang, I. S., Crouch, M. L., and Fang, F. C. (2005). Compensatory role of PspA, a member of the phage shock protein operon, in rpoE mutant Salmonella enterica serovar Typhimurium. Mol Microbiol *56*, 1004-1016.

Bejerano-Sagie, M., and Xavier, K. B. (2007). The role of small RNAs in quorum sensing. Curr Opin Microbiol *10*, 189-198.

Beloglazova, N., Brown, G., Zimmerman, M. D., Proudfoot, M., Makarova, K. S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W.*, et al.* (2008). A novel family of sequence-specific endoribonucleases associated

with the clustered regularly interspaced short palindromic repeats. J Biol Chem *283*, 20361-20371.

Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors? Mol Microbiol *22*, 393-404.

Berks, B. C., Palmer, T., and Sargent, F. (2003). The Tat protein translocation pathway and its role in microbial physiology. Adv Microb Physiol *47*, 187-254.

Berks, B. C., Palmer, T., and Sargent, F. (2005). Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. Curr Opin Microbiol *8*, 174-181.

Berks, B. C., Sargent, F., and Palmer, T. (2000). The Tat protein export pathway. Mol Microbiol *35*, 260-274.

Blaudeck, N., Kreutzenbeck, P., Freudl, R., and Sprenger, G. A. (2003). Genetic analysis of pathway specificity during posttranslational protein translocation across the Escherichia coli plasma membrane. J Bacteriol *185*, 2811-2819.

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.

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. Curr Opin Microbiol *10*, 102-109.

Brennan, R. G., and Link, T. M. (2007). Hfq structure, function and ligand binding. Curr Opin Microbiol *10*, 125-133.

Brissette, J. L., Russel, M., Weiner, L., and Model, P. (1990). Phage shock protein, a stress protein of Escherichia coli. Proc Natl Acad Sci U S A *87*, 862-866.

Bronstein, P. A., Marrichi, M., Cartinhour, S., Schneider, D. J., and DeLisa, M. P. (2005). Identification of a twin-arginine translocation system in Pseudomonas syringae pv. tomato DC3000 and its contribution to pathogenicity and fitness. J Bacteriol *187*, 8450-8461.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., 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.

Bukau, B. (1993). Regulation of *E. coli* heat shock response. Molecular Microbiology *9*, 671-680.

Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N.*, et al.* (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature *433*, 531-537.

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.

Castanie, M. P., Berges, H., Oreglia, J., Prere, 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.

Chanal, A., Santini, C., and Wu, L. (1998). Potential receptor function of three homologous components, TatA, TatB and TatE, of the twin-arginine signal sequence-dependent metalloenzyme translocation pathway in Escherichia coli. Mol Microbiol *30*, 674-676.

Chanal, A., Santini, C. L., and Wu, L. F. (2003). Specific inhibition of the translocation of a subset of Escherichia coli TAT substrates by the TorA signal peptide. J Mol Biol *327*, 563-570.

Chen, G., Hayhurst, A., Thomas, J. G., Harvey, B. R., Iverson, B. L., and Georgiou, G. (2001). Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). Nat Biotechnol *19*, 537-542.

Collier, D. N., Bankaitis, V. A., Weiss, J. B., and Bassford, P. J., Jr. (1988). The antifolding activity of SecB promotes the export of the E. coli maltose-binding protein. Cell *53*, 273-283.

Craig, J. E., Nobbs, A., and High, N. J. (2002). The extracytoplasmic sigma factor, final σE, is required for intracellular survival of nontypeable Haemophilus influenzae in J774 macrophages. Infect Immun *70*, 708 715.

Cristobal, S., de Gier, J. W., Nielsen, H., and von Heijne, G. (1999). Competition between Sec- and TAT-dependent protein translocation in Escherichia coli. Embo J *18*, 2982-2990.

Danese, P. N., and Silhavy, T. J. (1997). The sigma(E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in Escherichia coli. Genes Dev *11*, 1183-1193.

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 U S A *97*, 6640-6645.

de Fougerolles, A., Vornlocher, H. P., Maraganore, J., and Lieberman, J. (2007). Interfering with disease: a progress report on siRNA-based therapeutics. Nat Rev Drug Discov *6*, 443-453.

De Las, P., Connolly, L., and Gross, C. A. (1997). σE is an essential sigma factor in Escherichia coli. J Bacteriol *179*, 6862 6864.

De Lay, N., and Gottesman, S. (2009). The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. J Bacteriol *191*, 461-476.

De Leeuw, E., Porcelli, I., Sargent, F., Palmer, T., and Berks, B. C. (2001). Membrane interactions and self-association of the TatA and TatB components of the twin-arginine translocation pathway. FEBS Lett *506*, 143-148.

De Wulf, P., McGuire, A. M., Liu, X., and Lin, E. C. (2002). Genome-wide profiling of promoter recognition by the two-component response regulator CpxR-P in Escherichia coli. J Biol Chem *277*, 26652-26661.

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 U S A *100*, 6115-6120.

Derman, A. I., Prinz, W. A., Belin, D., and Beckwith, J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli. Science *262*, 1744-1747.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature *400*, 693-696.

Diao, J., Young, L., Kim, S., Fogarty, E. A., Heilman, S. M., Zhou, P., Shuler, M. L., Wu, M., and DeLisa, M. P. (2006). A three-channel microfluidic device for generating static linear gradients and its application to the quantitative analysis of bacterial chemotaxis. Lab Chip *6*, 381-388.

Dionisi, H. M., Checa, S. K., Krapp, A. R., Arakaki, A. K., Ceccarelli, E. A., Carrillo, N., and Viale, A. M. (1998). Cooperation of the DnaK and GroE chaperone systems in the folding pathway of plant ferredoxin-NADP+ reductase expressed in Escherichia coli. Eur J Biochem *251*, 724-728.

Driessen, A. J., Manting, E. H., and van der Does, C. (2001). The structural basis of protein targeting and translocation in bacteria. Nat Struct Biol *8*, 492-498.

Dubini, A., and Sargent, F. (2003). Assembly of Tat-dependent [NiFe] hydrogenases: identification of precursor-binding accessory proteins. FEBS Lett *549*, 141-146.

Duguay, A. R., and Silhavy, T. J. (2004). Quality control in the bacterial periplasm. Biochim Biophys Acta *1694*, 121-134.

Dykxhoorn, D. M., Palliser, D., and Lieberman, J. (2006). The silent treatment: siRNAs as small molecule drugs. Gene Ther *13*, 541-552.

Echeverri, C. J., and Perrimon, N. (2006). High-throughput RNAi screening in cultured cells: a user's guide. Nat Rev Genet *7*, 373-384.

Ehrmann, M., and Clausen, T. (2004). Proteolysis as a regulatory mechanism. Annu Rev Genet *38*, 709-724.

Enoch, H. G., and Lester, R. L. (1975). The purification and properties of formate dehydrogenase and nitrate reductase from Escherichia coli. J Biol Chem *250*, 6693-6705.

Erbse, A., Dougan, D. A., and Bukau, B. (2003). A folding machine for many but a master of none. Nat Struct Biol *10*, 84-86.

Farrell, C. M., Grossman, A. D., and Sauer, R. T. (2005). Cytoplasmic degradation of ssrA-tagged proteins. Mol Microbiol *57*, 1750-1761.

Fisher, A. C., and DeLisa, M. P. (2004). A little help from my friends: quality control of presecretory proteins in bacteria. J Bacteriol *186*, 7467-7473.

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.

Fozo, E. M., Hemm, M. R., and Storz, G. (2008). Small toxic proteins and the antisense RNAs that repress them. Microbiol Mol Biol Rev *72*, 579-589, Table of Contents.

Gammell, P., Barron, N., Kumar, N., and Clynes, M. (2007). Initial identification of low temperature and culture stage induction of miRNA expression in suspension CHO-K1 cells. J Biotechnol *130*, 213-218.

Gon, S., Patte, J. C., Mejean, V., and Iobbi-Nivol, C. (2000). The torYZ (yecK bisZ) operon encodes a third respiratory trimethylamine N-oxide reductase in Escherichia coli. J Bacteriol *182*, 5779-5786.

Gottesman, S. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet *21*, 399-404.

Govan, J. R., Martin, D. W., and Deretic, V. P. (1992). Mucoid Pseudomonas aeruginosa and cystic fibrosis: the role of mutations in muc loci. FEMS Microbiol Lett *79*, 323-329.

Grissa, I., Vergnaud, G., and Pourcel, C. (2008). CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. Nucleic Acids Res *36*, W145-148.

Groenen, P. M., Bunschoten, A. E., van Soolingen, D., and van Embden, J. D. (1993). Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. Mol Microbiol *10*, 1057-1065.

Guerrier-Takada, C., Li, Y., and Altman, S. (1995). Artificial regulation of gene expression in Escherichia coli by RNase P. Proc Natl Acad Sci U S A *92*, 11115-11119.

Gur, E., Biran, D., Shechter, N., Genevaux, P., Georgopoulos, C., and Ron, E. Z. (2004). The Escherichia coli DjlA and CbpA proteins can substitute for DnaJ in DnaK-mediated protein disaggregation. J Bacteriol *186*, 7236-7242.

Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol *177*, 4121-4130.

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 Comput Biol *1*, e60.

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

Harmon, K. S., and McKay, L. L. (1987). Restriction enzyme analysis of lactose and bacteriocin plasmids from Streptococcus lactis subsp. diacetylactis WM4 and cloning of BclI fragments coding for bacteriocin production. Appl Environ Microbiol *53*, 1171-1174.

Harth, G., Zamecnik, P. C., Tabatadze, D., Pierson, K., and Horwitz, M. A. (2007). Hairpin extensions enhance the efficacy of mycolyl transferase-specific antisense oligonucleotides targeting Mycobacterium tuberculosis. Proc Natl Acad Sci U S A *104*, 7199-7204.

Hebert, C. G., Valdes, J. J., and Bentley, W. E. (2008). Beyond silencing--engineering applications of RNA interference and antisense technology for altering cellular phenotype. Curr Opin Biotechnol *19*, 500-505.

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.

Horvath, P., Barrangou, R., Fremaux, C., Boyaval, P., and Romero, D. (2007). International Patent Application.

Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F., and Hartl, F. U. (1999). Identification of in vivo substrates of the chaperonin GroEL. Nature *402*, 147-154.

Howell, J. M., and Turner, R. J. (2006). Identification of an interactome for DmsD, a chaperone for DmsA and the Tat system. J Biol Chem (submitted).

Humphreys, S., Stevenson, A., Bacon, A., Weinhardt, A. B., and Roberts, M. T. (1999). The alternative sigma factor, σE, is critically important for the virulence of Salmonella typhimurium. Infect Immun *67*, 1560 1568.

Jack, R. L., Buchanan, G., Dubini, A., Hatzixanthis, K., Palmer, T., and Sargent, F. (2004). Coordinating assembly and export of complex bacterial proteins. Embo J *23*, 3962-3972.

Jansen, R., Embden, J. D., Gaastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565-1575.

Jones, S. E., Lloyd, L. J., Tan, K. K., and Buck, M. (2003). Secretion defects that activate the phage shock response of Escherichia coli. J Bacteriol *185*, 6707–6711.

Jong, W. S., ten Hagen-Jongman, C. M., Genevaux, P., Brunner, J., Oudega, B., and Luirink, J. (2004). Trigger factor interacts with the signal peptide of nascent Tat substrates but does not play a critical role in Tat-mediated export. Eur J Biochem *271*, 4779-4787.

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., and van Embden, J. (1997). Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol *35*, 907-914.

Kawano, M., Aravind, L., and Storz, G. (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Mol Microbiol *64*, 738-754.

Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H. C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., and Hartl, F. U. (2005). Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell *122*, 209-220.

Kim, J.-Y., Fogarty, E. A., Lu, F. J., Zhu, H., 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.

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.

Kleerebezem, M., Crielaard, W., and Tommassen, J. (1996). Involvement of stress protein PspA (phage shock protein A) of Escherichia coli in maintenance of the protonmotive force under stress conditions. Embo J *15*, 162-171.

Kleerebezem, M., and Tommassen, J. (1993). Expression of the pspA gene stimulates efficient protein export in Escherichia coli. Mol Microbiol *7*, 947-956.

Knoblauch, N. T., Rudiger, S., Schonfeld, H. J., Driessen, A. J., Schneider-Mergener, J., and Bukau, B. (1999). Substrate specificity of the SecB chaperone. J Biol Chem *274*, 34219-34225.

Koop, A. H., Hartley, M. E., and Bourgeois, S. (1987). A low-copy-number vector utilizing beta-galactosidase for the analysis of gene control elements. Gene *52*, 245-256.

Kumamoto, C. A. (1989). Escherichia coli SecB protein associates with exported protein precursors in vivo. Proc Natl Acad Sci U S A *86*, 5320-5324.

Kumamoto, C. A., and Beckwith, J. (1983). Mutations in a new gene, secB, cause defective protein localization in Escherichia coli. J Bacteriol *154*, 253-260.

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

Kunst, F., Debarbouille, M., Msadek, T., Young, M., Mauel, C., Karamata, D., Klier, A., Rapoport, G., and Dedonder, R. (1988). Deduced polypeptides encoded by the Bacillus subtilis sacU locus share homology with two-component sensor-regulator systems. J Bacteriol *170*, 5093-5101.

Lai, D., Fu, L., Weng, S., Qi, L., Yu, C., Yu, T., Wang, H., and Chen, W. (2004). Blocking caspase-3 activity with a U6 SnRNA promoter-driven ribozyme enhances survivability of CHO cells cultured in low serum medium and production of interferon-beta. Biotechnol Bioeng *85*, 20-28.

Lecker, S. H., Driessen, A. J., and Wickner, W. (1990). ProOmpA contains secondary and tertiary structure prior to translocation and is shielded from aggregation by association with SecB protein. Embo J *9*, 2309-2314.

Leheny, E. A., Teter, S. A., and Theg, S. M. (1998). Identification of a Role for an Azide-Sensitive Factor in the Thylakoid Transport of the 17-Kilodalton Subunit of the Photosynthetic Oxygen-Evolving Complex. Plant Physiol *116*, 805-814.

Li, S. Y., Chang, B. Y., and Lin, S. C. (2006). Coexpression of TorD enhances the transport of GFP via the TAT pathway. J Biotechnol *122*, 412-421.

Lillestol, R. K., Redder, P., Garrett, R. A., and Brugger, K. (2006). A putative viral defence mechanism in archaeal cells. Archaea *2*, 59-72.

Liu, G., Topping, T. B., and Randall, L. L. (1989). Physiological role during export for the retardation of folding by the leader peptide of maltose-binding protein. Proc Natl Acad Sci U S A *86*, 9213-9217.

Loosmore, S. M., Yang, Y. P., Oomen, R., Shortreed, J. M., Coleman, D. C., and Klein, M. H. (1998). The Haemophilus influenzae HtrA protein is a protective antigen. Infect Immun *66*, 899-906.

Lopez-Sanchez, F., Ramirez-Santos, J., and Gomez-Eichelmann, M. (1997). *In vivo* effect of DNA relaxation on the transcription of gene *rpoH* in *Escherichia coli*. Biochem Biophys Acta *1353*, 79-83.

Ma, E., MacRae, I. J., Kirsch, J. F., and Doudna, J. A. (2008). Autoinhibition of human dicer by its internal helicase domain. J Mol Biol *380*, 237-243.

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.

Madueno, F., Napier, J. A., and Gray, J. C. (1993). Newly Imported Rieske Iron-Sulfur Protein Associates with Both Cpn60 and Hsp70 in the Chloroplast Stroma. Plant Cell *5*, 1865-1876.

Majors, B. S., Betenbaugh, M. J., and Chiang, G. G. (2007). Links between metabolism and apoptosis in mammalian cells: applications for anti-apoptosis engineering. Metab Eng *9*, 317-326.

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.

March, J. C., and Bentley, W. E. (2006). Engineering eukaryotic signal transduction with RNAi: enhancing Drosophila S2 cell growth and recombinant protein synthesis via silencing of TSC1. Biotechnol Bioeng *95*, 645-652.

March, J. C., and Bentley, W. E. (2007). RNAi-based tuning of cell cycling in Drosophila S2 cells--effects on recombinant protein yield. Appl Microbiol Biotechnol *73*, 1128-1135.

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

Meng, J., Hu, B., Liu, J., Hou, Z., Meng, J., Jia, M., and Luo, X. (2006). Restoration of oxacillin susceptibility in methicillin-resistant Staphylococcus aureus by blocking the MecR1-mediated signaling pathway. J Chemother *18*, 360-365.

Mikhaleva, N. I., Santini, C. L., Giordano, G., Nesmeyanova, M. A., and Wu, L. F. (1999). Requirement for phospholipids of the translocation of the trimethylamine N-oxide reductase through the Tat pathway in Escherichia coli. FEBS Lett *463*, 331-335.

Miticka, H., Rowley, G., Rezuchova, B., Homerova, D., Humphreys, S., Farn, J., Roberts, M., and Kormanec, J. (2003). Transcriptional analysis of the rpoE gene encoding extracytoplasmic stress response sigma factor sigmaE in Salmonella enterica serovar Typhimurium. FEMS Microbiol Lett *226*, 307-314.

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

Mogk, A., Deuerling, E., Vorderwulbecke, S., Vierling, E., and Bukau, B. (2003). Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. Mol Microbiol *50*, 585-595.

Mojica, F. J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol *36*, 244-246.

Molik, S., Karnauchov, I., Weidlich, C., Herrmann, R. G., and Klosgen, R. B. (2001). The Rieske Fe/S protein of the cytochrome b6/f complex in

chloroplasts: missing link in the evolution of protein transport pathways in chloroplasts? J Biol Chem *276*, 42761-42766.

Montange, R. K., and Batey, R. T. (2008). Riboswitches: emerging themes in RNA structure and function. Annu Rev Biophys *37*, 117-133.

Nevesinjac, A. Z., and Raivio, T. L. (2005). The Cpx envelope stress response affects expression of the type IV bundle-forming pili of enteropathogenic Escherichia coli. J Bacteriol *187*, 672-686.

Nudler, E., and Mironov, A. S. (2004). The riboswitch control of bacterial metabolism. Trends Biochem Sci *29*, 11-17.

Nyfeler, B., Michnick, S. W., and Hauri, H. P. (2005). Capturing protein interactions in the secretory pathway of living cells. Proc Natl Acad Sci U S A *102*, 6350-6355.

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

Palmer, T., Sargent, F., and Berks, B. C. (2005). Export of complex cofactor-containing proteins by the bacterial Tat pathway. Trends Microbiol *13*, 175-180.

Perez-Rodriguez, 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.

Phillips, G. J., and Silhavy, T. J. (1990). Heat-shock proteins DnaK and GroEL facilitate export of LacZ hybrid proteins in E. coli. Nature *344*, 882-884.

Pogliano, J., Lynch, A. S., Belin, D., Lin, E. C., and Beckwith, J. (1997). Regulation of Escherichia coli cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. Genes Dev *11*, 1169-1182.

Prevost, K., Salvail, H., Desnoyers, G., Jacques, J. F., Phaneuf, E., and Masse, E. (2007). The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. Mol Microbiol *64*, 1260-1273.

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

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

Raivio, T., and Silhavy, T. (2001). Periplasmic Stress and ECF sigma factors. Annual Revisions in Microbiology *55*, 591-624.

Raivio, T. L., and Silhavy, T. J. (1999). The sigmaE and Cpx regulatory pathways: overlapping but distinct envelope stress responses. Curr Opin Microbiol *2*, 159-165.

Rao, R. V., and Bredesen, D. E. (2004). Misfolded proteins, endoplasmic reticulum stress and neurodegeneration. Curr Opin Cell Biol *16*, 653-662.

Ray, N., Oates, J., Turner, R. J., and Robinson, C. (2003). DmsD is required for the biogenesis of DMSO reductase in Escherichia coli but not for the interaction of the DmsA signal peptide with the Tat apparatus. FEBS Lett *534*, 156-160.

Rezuchova, B., Miticka, H., Homerova, D., Roberts, M., and Kormanec, J. (2003). New members of the Escherichia coli σE regulon identified by a two-plasmid system. FEMS Microbiol Lett *225*, 1-7.

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

Robinson, C., and Bolhuis, A. (2001). Protein targeting by the twin-arginine translocation pathway. Nat Rev Mol Cell Biol *2*, 350-356.

Rodrigue, A., Batia, N., Muller, M., Fayet, O., Bohm, R., Mandrand-Berthelot, M. A., and Wu, L. F. (1996). Involvement of the GroE chaperonins in the nickel-dependent anaerobic biosynthesis of NiFe-hydrogenases of Escherichia coli. J Bacteriol *178*, 4453-4460.

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

Rowley, G., Spector, M., Kormanec, J., and Roberts, M. (2006). Pushing the envelope: extracytoplamic stress responses in bacterial pathogens. Nature reviews, Microbiology *4*.

Rudiger, S., Germeroth, L., Schneider-Mergener, J., and Bukau, B. (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. Embo J *16*, 1501-1507.

Ruiz, N., Kahne, D., and Silhavy, T. J. (2006). Advances in understanding bacterial outer-membrane biogenesis. Nature Rev Microbiol *4*, 57-66.

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.

Santini, C. L., Bernadac, A., Zhang, M., Chanal, A., Ize, B., Blanco, C., and Wu, L. F. (2001). Translocation of jellyfish green fluorescent protein via the Tat system of Escherichia coli and change of its periplasmic localization in response to osmotic up-shock. J Biol Chem *276*, 8159-8164.

Sargent, F., Bogsch, E. G., Stanley, N. R., Wexler, M., Robinson, C., Berks, B. C., and Palmer, T. (1998). Overlapping functions of components of a bacterial Sec-independent protein export pathway. EMBO J *17*, 3640-3650.

Silvestro, A., Pommier, J., and Giordano, G. (1988). The inducible trimethylamine-N-oxide reductase of Escherichia coli K12: biochemical and immunological studies. Biochim Biophys Acta *954*, 1-13.

Soler Bistue, A. J., Ha, H., Sarno, R., Don, M., Zorreguieta, A., and Tolmasky, M. E. (2007). External guide sequences targeting the aac(6')-Ib mRNA induce inhibition of amikacin resistance. Antimicrob Agents Chemother *51*, 1918-1925.

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.

Stanley, N. R., Findlay, K., Berks, B. C., and Palmer, T. (2001). Escherichia coli strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. J Bacteriol *183*, 139-144.

Stephanopoulos, G., and Vallino, J. J. (1991). Network rigidity and metabolic engineering in metabolite overproduction. Science *252*, 1675-1681.

Takiff, H. E., Chen, S. M., and Court, D. L. (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 Huttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. Proc Natl Acad Sci U S A *99*, 7536-7541.

Tatsuta, T., Tomoyasu, T., Bakau, B., Kitagawa, M., Mori, H., Karata, K., and Ogura, T. (1998). Heat shock regulation in the *ftsH* null mutant of *Escherichia coli*: dissection of stability and activity control mechanisms of sigma32 in vivo. Molecular Microbiology *30*, 583-593.

Thomas, J., and Baneyx, B. (1998). Roles of *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG *in vivo*. Journal of bacteriology *180*, 5165--5172.

Thomas, J. D., Daniel, R. A., Errington, J., and Robinson, C. (2001). Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in Escherichia coli. Mol Microbiol *39*, 47-53.

Tomoyasu, T., Ogura, T., Tatsuta, T., and Bukau, B. (1998). Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in Escherichia coli. Mol Microbiol *30*, 567-581.

Trotochaud, A. E., and Wassarman, K. M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. Nat Struct Mol Biol *12*, 313-319.

Tullman-Ercek, D., DeLisa, M. P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., and Georgiou, G. (2006). Export pathway selectivity of *Escherichia coli* twin-arginine translocation signal peptides. (submitted).

Tullman-Ercek, D., DeLisa, M. P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., and Georgiou, G. (2007). Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides. J Biol Chem *282*, 8309-8316.

Turner, R. J., Papish, A. L., and Sargent, F. (2004). Sequence analysis of bacterial redox enzyme maturation proteins (REMPs). Can J Microbiol *50*, 225-238.

Van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., and Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36-44.

Van Dyk, T., Majarian, W., Konstanstinov, K., Young, R., Dhurjati, P., and LaRossa, R. (1994). Rapid and Sensitive Pollutant Detection by Induction of Heat Shock Gene-Bioluminescence Gene Fusions. Applied and Environmental Microbiology *60*, 1414-1420.

Vickery, L., Silberg, J., and Ta, D. (1997). Hsc66 and Hsc20, a new heat shock cognate molecular chaperone system from Escherichia coli. Protein Science *6*, 1047-1056.

Vogel, J., Argaman, L., Wagner, E. G., and Altuvia, S. (2004). The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. Curr Biol *14*, 2271-2276.

Wagner, E. G., Altuvia, S., and Romby, P. (2002). Antisense RNAs in bacteria and their genetic elements. Adv Genet *46*, 361-398.

Watanabe, M., and Blobel, G. (1989). Cytosolic factor purified from Escherichia coli is necessary and sufficient for the export of a preprotein and is a homotetramer of SecB. Proc Natl Acad Sci U S A *86*, 2728-2732.

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

Weiner, L., and Model, P. (1994). Role of an Escherichia coli stress-response operon in stationary-phase survival. Proc Natl Acad Sci U S A *91*, 2191-2195.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., and Doudna, J. A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure *17*, 904-912.

Wild, J., Altman, E., Yura, T., and Gross, C. A. (1992). DnaK and DnaJ heat shock proteins participate in protein export in Escherichia coli. Genes Dev *6*, 1165-1172.

Wild, J., Rossmeissl, P., Walter, W. A., and Gross, C. A. (1996). Involvement of the DnaK-DnaJ-GrpE chaperone team in protein secretion in Escherichia coli. J Bacteriol *178*, 3608-3613.

Wong, D. C., Wong, K. T., Nissom, P. M., Heng, C. K., and Yap, M. G. (2006). Targeting early apoptotic genes in batch and fed-batch CHO cell cultures. Biotechnol Bioeng *95*, 350-361.

Xu, Y., Weng, C. L., Narayanan, N., Hsieh, M. Y., Anderson, W. A., Scharer, J. M., Moo-Young, M., and Chou, C. P. (2005). Chaperone-mediated folding and maturation of the penicillin acylase precursor in the cytoplasm of Escherichia coli. Appl Environ Microbiol *71*, 6247-6253.

Yahr, T. L., and Wickner, W. T. (2001). Functional reconstitution of bacterial Tat translocation in vitro. Embo J *20*, 2472-2479.

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.

Yang, S. W., Xu, L., Mierzwa, R., He, L., Terracciano, J., Patel, M., Gullo, V., Black, T., Zhao, W., Chan, T. M., and Chu, M. (2004). Two novel antibiotics, Sch 419558 and Sch 419559, produced by Pseudomonas fluorescens: effect on activity by overexpression of RpoE. Bioorg Med Chem *12*, 3333-3338.

Yura, T. (1993). Regulation of heat response in bacteria. Annual Revisions in Microbiology *47*, 321-350.

Yura, T., and Nakahigashi, K. (1999). Regulation of the heat shock response. Current Opinion in Microbiology *2*, 153-158.

Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G., and Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol Microbiol *50*, 1111-1124.