Title
Investigating the Determinants of Ribosome Pausing in Bacteria and Cellular Function of Trigger Factor

Author
Oh, Eugene

Publication Date
2012

Peer reviewed|Thesis/dissertation

Investigating the Determinants of Ribosome Pausing in Bacteria and Cellular Function of Trigger Factor

by

Eugene Oh

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

ACKNOWLEDGEMENTS

My success as a researcher could not have been possible without the generosity, patience, and support of other scientists. As I reflect over the past few years, I have many people to thank for shaping my scientific trajectory.

The Weissman lab is filled with some of the most brilliant and talented people who somehow remain incredibly humble and grounded. I would like to thank Mike Bassik, Gloria Brar, Clement Chu, Nick Ingolia, Calvin Jan, Gene-Wei Li, Martin Kampmann, Noam Stern-Ginossar, and Jimena Weibezahn for all their guidance. I would also like to the thank my student compatriots Liz Boydston, Dave Breslow, Josh Dunn, Cat Foo, Martin Jonikas, Yi-Chang Liu, Edwin Rodriguez, Silvi Rouskin, Mel Smith, Jacob Stewart-Ornstein, and Chris Williams, who all provided a fun and engaging lab environment.

My thesis work could not have been possible without close collaborations, specifically those from the labs of Bernd Bukau and Carol Gross. I could not have wished for a better colleague than Annemarie Becker. I admire her determination and intelligence, as well as her generosity and patience. I also extend my sincere gratitude to Felix Gloge, Damon Huber, Günter Kramer, and Arzu Sandicki for work that ultimately enabled the publication of our manuscript. In addition, my foray into bacterial genetics and cell biology could not have been possible without the help of David Burkhardt, Rachna Chaba, Andrew Gray, Bentley Lim, Bob Nichols, and Nassos Typas.

I doubt that I could have been the scientist I am today without the tutelage of Jonathan Weissman. I am grateful he took a chance on me. Jonathan is on a level so far removed from the rest of us that all you can do is marvel in awe. He is remarkable in so many ways, but I would like to highlight his ingenuity, intuition, and generosity. One of

my fondest experiences was at a scientific meeting I attended with Jonathan, Bernd, and Carol. Jonathan was going to present later that night, and we sat around crafting how he should deliver our newest findings. It was intense, yet so much fun. To be included as peer in the likes of these heavyweights may have been the most rewarding part of my graduate career.

I would also like to thank John Abelson, Raul Andino, Carol Gross, and Sandy Johnson for sitting on various committees. In addition, I would like to thank Manny DeVera, Christopher Reiger, and Alice Willis-Breezi for making the lab function, as well as Sue Adams, Danny Dam, and Rachel Mozesson for running the graduate program.

Finally, I would like to acknowledge my former mentor Jon Widom, who passed during my graduate career. Even though our paths crossed briefly, he will be sorely missed for his creativity and brilliance. His contributions to the field of chromatin biology, specifically on how DNA is packaged, have had far reaching implications on how genes are expressed in cells. His resolve to apply physical principles to complex nature of biological systems made me appreciate math once more.

ABSTRACT

Biological systems use proteins to drive just about every cellular process. Thus one would gain much insight into the physiology of an organism by knowing the catalog of proteins that are actively expressed. Ribosome profiling (deep sequencing of ribosome-protected mRNA footprints) enables one to monitor protein production genome-wide. Here, I describe the application of ribosome profiling in the Gram-negative bacterium *Escherichia* coli, and show that translation control is pervasive. In addition, I review the regulatory mechanisms that govern transient ribosome pausing, which can influence a number of cotranslational processes, such as folding and localization of the nascent chain. I continue this discussion by showing that pausing is widespread, and mainly stems from transient interactions made between the anti-Shine-Dalgarno sequence of the 16S ribosomal RNA and Shine-Dalgarno-like sequences contained within messenger RNA, rather than being determined by the concentrations of transfer RNA. Finally, I describe a new application of the ribosome profiling approach that enables one to monitor cotranslational processes in vivo. Here, I characterize the cotranslational action of the molecular chaperone trigger factor. This study unexpectedly revealed that trigger factor does not engage the nascent chain until well after its synthesis, contrary to findings from in vitro experiments, and that outer membrane porin proteins are the most prominent substrates of this chaperone.

CONTRIBUTIONS

Excluding the contributions listed below, work presented in this dissertation was performed by Eugene Oh. Chapter III, entitled “The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria,” first appeared in print on April 26, 2012 in Nature, *484*(7395), 538–541. From my preliminary data, Dr. Gene-Wei Li discovered that transient elongation arrest is mediated by the anti-Shine-Dalgarno sequence present in ribosomes, and not by cellular levels of tRNA. He also showed that internal Shine-Dalgano sequences are disfavored in protein coding sequences, leading to their biased usage. This work was supervised entirely by Dr. Jonathan S. Weissman. Chapter IV, entitled “Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo,” first appeared in print on December 9, 2011 in Cell, *147*(6), 1295–1308, and is reproduced with permission from Elsevier. This work was performed in close collaboration with Annemarie H. Becker of the University of Heidelberg, Germany. Ms. Becker established the crosslinking protocol necessary for stabilizing interactions between trigger factor and ribosome-nascent chain complexes. In addition, she independently quantified proteins isolated from outer membranes by mass spectrometry. Arzu Sandikci and Felix Golge carried out assays showing the interplay between trigger factor and the processing enzymes PDF and MAP. Dr. Damon Huber performed pulse-labeling experiments showing the precise mode of translocation for specific substrates. Dr. Rachna Chaba measured in vivo levels of σ^E activity. Altogether, this work was supervised under Drs. Bernd Bukau and Jonathan S. Weissman.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER THREE

LIST OF FIGURES, CONTINUED

CHAPTER FOUR

CHAPTER ONE

Introduction

Background

Cellular processes are largely driven by the concerted action of proteins. As a result, proteins are the most abundant class of biological macromolecules. Additionally, proteins perform diverse roles, serving as, for example, antibodies, hormones, reaction catalysts, and structural scaffolds. Likewise, proteins vary in composition, size, and structure. Remarkably, cells can synthesize a near infinite combination of proteins using a finite number of building blocks. In this process, the ribosome decodes the sequence of codons on an mRNA template (Ramakrishnan, 2002), properly orienting matched aminoacylated tRNAs to catalyze the formation of peptide bonds (Rodnina et al., 2007).

Bacterial ribosomes can synthesize ~20 amino acids per second (Frank, 2003). Yet the rate of translation is not uniform across a given message (Li et al., 2012). In fact, transient pausing can regulate protein production (Gong and Yanofsky, 2002; Nakatogawa and Ito, 2002), folding (Kimchi-Sarfaty et al., 2007; Zhang et al., 2009), and localization of the polypeptide (Mariappan et al., 2010) or mRNA (Yanagitani et al., 2011). Such pauses can result from codon usage (Sørensen et al., 1989), mRNA structure (Namy et al., 2006), or the nascent chain (Cruz-Vera et al., 2011). In Chapter II, I review how cells exploit chain-induced translation arrest to regulate many aspects of gene expression. I carry on this discussion in Chapter III by addressing what drives transient elongation arrest in bacterial cells.

Knowing the collection of proteins that are actively made can yield much insight into any physiological process (Brar et al., 2012; Ingolia et al., 2009). By extension, the systematic characterization of nascent chains that are modified, folded, and targeted can reveal interesting new rules that govern cotranslational processes (Del Alamo et al., 2011; Oh et al., 2011). In Chapter IV, I describe an approach for characterizing bacterial translation in vivo and monitoring the cotranslational chaperone action of trigger factor.

The Biology of Trigger Factor

Denatured polypeptides can fold in a matter of seconds, with hydrophobic residues being buried inside the overall structure (Hartl and Hayer-Hartl, 2009). In the cell, spontaneous folding would be aggregation-prone and time-consuming given the crowded environment (Zimmerman and Trach, 1991). Additionally, proteases may degrade extended chains before folding into resistant forms (Hoffmann et al., 2010). But ultimately, folding is intimately coupled to translation (Hartl and Hayer-Hartl, 2009; Kramer et al., 2009). So nascent chains may begin the folding process well before the full-length polypeptide is made. Moreover, C-terminal regions may be required to establish the native fold. For these reasons, protein folding in vivo is far more complex (Hartl and Hayer-Hartl, 2009). However, cells have evolved molecular chaperones that increase the capacity of folding (Kramer et al., 2009; Preissler and Deuerling, 2012). For example, chaperones can shield nascent chains from aggregation, actively fold polypeptides, assemble higher-order complexes, and disengage aggregates.

In general, molecular chaperones can be classified into two groups (Hoffmann et al., 2010), those that engage the nascent chain cotranslationally versus those that act on the mature polypeptide. In bacteria, trigger factor (TF) is the sole ribosome-associated chaperone (Hoffmann et al., 2010). TF functions by protecting nascent chains from aggregation (Merz et al., 2006) and protease-mediated degradation (Hoffmann et al., 2006; Tomic et al., 2006). In addition, TF may improve folding efficiency (i.e. specific activity of the polypeptide) by slowing down the speed of folding (Agashe et al., 2004). By contrast, eukaryotes use two mechanistically distinct ribosome-associated chaperone systems (Preissler and Deuerling, 2012), an Hsp70/40-based triad (Gautschi et al., 2001; Hundley et al., 2005) and the nascent chain-associated complex (NAC) (Rospert et al., 2002). In yeast, Ssb is the Hsp70 homologue (Peisker et al., 2010) whose ATPase

activity is modulated by the ribosome-associated complex (RAC) (Gautschi et al., 2001). Cells lacking Ssb or RAC subunits show similar growth defects when under folding stress (Gautschi et al., 2002), indicating they are functionally related. It is presently unknown how NAC chaperones the emerging nascent chain, but cells lacking both Ssb and NAC lose viability upon treatment with antibiotics that exacerbate the folding load (Koplin et al., 2010), ultimately driving terminal aggregation. Remarkably, TF can partially rescue the aminoglycoside sensitivity observed in cells lacking the Ssb-RAC chaperone system (Rauch et al., 2005), even though they are unrelated mechanistically.

In 1987, Bill Wickner discovered TF through a biochemical screen (Crooke and Wickner, 1987). When TF was added during the dialysis (or refolding) of denatured pro-OmpA, the refolded chains could translocate across inside-out vesicles. Interestingly, when TF was added after the refolding of denatured pro-OmpA, they were no longer competent for membrane insertion, suggesting that TF interacts with the unfolded form for function. Surprisingly, pro-OmpA chains refolded by rapid dilution were competent for secretion even without the aid of TF (Crooke et al., 1988). This apparent contradiction was resolved when it was demonstrated that rapidly refolded pro-OmpA loses the ability to translocate over time (Crooke et al., 1988), indicating that TF maintains the secretion-competent form. Indeed, pro-OmpA chains refolded by rapid dilution were stabilized by the presence of TF over long incubations (Crooke et al., 1988). Nonetheless, the physiological role of TF came into doubt when cells lacking TF did not show an obvious secretion defect (Guthrie and Wickner, 1990). Instead, Δ*tig* cells were hyper-filamentous in morphology, an effect that was rescued by the overexpression of FtsZ, which regulates cell division (Guthrie and Wickner, 1990).

Not long after its discovery, it was discovered that TF interacts with translating ribosomes (Lill et al., 1988) and those stalled by chloramphenicol (Hesterkamp et al.,

1996). Moreover, this interaction could be disrupted by puromycin (Hesterkamp et al., 1996), hinting that TF associates with the nascent chain. Indeed, TF can crosslink to model chains of stalled ribosome complexes (Hesterkamp et al., 1996; Valent et al., 1995). Additionally, TF was found to be one of the most potent peptidyl prolyl *cis*/trans isomerases in vitro (Hesterkamp et al., 1996; Scholz et al., 1997; Stoller et al., 1995), suggesting that proline isomerization may initiate on the ribosome, but this remains to be verified.

Over a decade after its discovery, TF was found to be essential in cells also lacking a secondary chaperone system, DnaKJ (Deuerling et al., 1999; Teter et al., 1999), hinting that TF may function similarly. Subsequent work showed that Δ*tig* cells induce the heat shock response by upregulating DnaKJ, as well as the chaperonin GroEL (Deuerling et al., 2003), indicating that cells lacking TF express compensatory chaperones to alleviate the increased folding load. Moreover, peptide scanning libraries revealed that ~80% of TF substrates are also recognized by DnaK (Deuerling et al., 2003; Patzelt et al., 2001; Rüdiger et al., 1997), validating that chaperones have redundant substrate pools, thus being better equipped to buffer one another when under folding stress. Accordingly, Δ*tig*Δ*dnaKJ* cells can be rescued by overexpressing GroEL (Vorderwülbecke et al., 2004). More recently, it has been demonstrated that Δ*tig*Δ*dnaKJ* cells can survive between 20-25ºC (Genevaux et al., 2004), reflecting a narrow range in which folding stress may be less pronounced. Interestingly, this phenotype stems from the irreversible aggregation of cytoplasmic proteins (~10% of the total) (Deuerling et al., 2003; Deuerling et al., 1999; Vorderwülbecke et al., 2004), but not membrane-associated ones, such as AtpD and OmpA (Deuerling et al., 1999). Altogether, these studies have affirmed the chaperone function of TF.

The association between TF and the ribosome has been extensively

characterized in vitro (Kaiser et al., 2006; Lakshmipathy et al., 2007; Maier et al., 2003; Merz et al., 2008; Patzelt et al., 2002; Raine et al., 2006; Rutkowska et al., 2008). It is assumed that translating ribosomes are always bound by TF, since their concentrations in vivo ([TF] = 20 μM; [70S] = 50 μM) (Lill et al., 1988) greatly exceed their affinity in vitro (K_D ~1–2 μM) (Maier et al., 2003; Patzelt et al., 2002). Specifically, this interaction is bridged by the ribosomal protein L23 (Kramer et al., 2002). Interestingly, the mutant variant of L23 incapable of binding TF loses viability when introduced to cells lacking DnaK (Kramer et al., 2002), stemming from toxicity toward the accumulation of aggregates, much like the phenotype observed in Δ*tig*Δ*dnaK* cells (see above).

SecB chaperones full-length presecretory proteins, ultimately targeting them to the translocon (Bernstein, 2000). As a result, Δ*secB* cells accumulate unprocessed chains in the cytoplasm (Lee and Bernstein, 2001). Yet surprisingly, the deletion of TF suppresses the export defect observed in Δ*secB* cells (Lee and Bernstein, 2002). Moreover, Δ*tig* cells accelerate the export of model substrates, while those overexpressing TF delay their secretion by sequestering them in the cytoplasm (Lee and Bernstein, 2002). A second group validated these findings in vivo (Ullers et al., 2007; Ullers et al., 2004). Accordingly, the deletion of TF suppresses the cold sensitivity seen in Δ*secB* cells, as well as the accumulation of toxic aggregates under standard growth conditions (Ullers et al., 2007). Additionally, increased levels of SecA and ribosomes associate with the inner membrane of Δ*tig* cells, while those ovexpressing TF accumulate modest amounts (Ullers et al., 2007). Altogether, these studies hint that TF could affect the cotranslational targeting of presecretory substrates. Interestingly, Δ*secB*Δ*dnaKJ* cells are inviable only in the presence of TF (Ullers et al., 2007), suggesting that the DnaKJ chaperone system can buffer SecB substrates (Sakr et al., 2010). It is presently unknown why the triple deletion remains viable. Nonetheless, I

elaborate on the in vivo consequences of losing TF function in Chapter IV.

Strikingly, cells lose viability when TF is induced ~4-fold more than endogenous levels (Genevaux et al., 2004). This effect is partly explained by the cytoplasmic accumulation of pro-OmpF chains, since Δ*ompF* cells tolerate up to ~10-fold more TF than wild-type cells (Genevaux et al., 2004). Thus the overexpression of TF delays the export of presecretory chains, ultimately inducing the terminal misfolding of essential factors in the cytoplasm.

When does TF interact with the nascent chain? Are certain sequences preferentially bound? Insight into how TF engages translating ribosomes have largely come from in vitro binding assays. TF can interact with vacant ribosomes, but this association is transient, lasting ~10–15 seconds (Kaiser et al., 2006; Maier et al., 2003; Patzelt et al., 2002). Depending on the length and sequence hydrophobicity, nascent chains accelerate TF association with L23, thereby increasing the overall affinity ~2–30-fold (Raine et al., 2006; Rutkowska et al., 2008). Thus nascent chains may control the timing of when TF binds to the ribosome. Following ribosome dissociation, TF can stay associated with the nascent chain for up to ~35 seconds (Kaiser et al., 2006), suggesting that multiple TF molecules can gain entry for chaperoning long multidomain proteins. But whether these findings can be recapitulated in vivo remains to be verified, and is the subject of Chapter IV.

A ribosome-independent role for TF was recently proposed by Wayne Hendrickson (Martinez-Hackert and Hendrickson, 2009). In this work, they showed that TF can stably associate with ~70 full-length proteins. Additionally, they solved the structure of tmTF (from *Thermotoga maritima*) in complex with tmS7 (one of the stably associated proteins), revealing that tmS7 assumes a near native fold, with tmTF masking the surface of tmS7 that is buried on assembly into the ribosome. From this model, it

was reasoned that TF may stabilize native-like structures before assembling them into higher-order complexes. Yet only a modest defect in 30S subunit assembly was seen in Δ*tig* cells (Martinez-Hackert and Hendrickson, 2009), casting doubt on the physiological relevance of this finding. Nonetheless, this work raises an important issue, which is also discussed in Chapter IV, addressing the unexplored functions of TF.

REFERENCES

Agashe, V.R., Guha, S., Chang, H.-C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.U., and Barral, J.M. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. Cell *117*, 199–209.

Bernstein, H.D. (2000). The biogenesis and assembly of bacterial membrane proteins. Curr. Opin. Microbiol. *3*, 203–209.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science *335*, 552–557.

Crooke, E., Brundage, L., Rice, M., and Wickner, W. (1988). ProOmpA spontaneously folds in a membrane assembly competent state which trigger factor stabilizes. EMBO J. *7*, 1831–1835.

Crooke, E., and Wickner, W. (1987). Trigger factor: a soluble protein that folds pro-OmpA into a membrane-assembly-competent form. Proc. Natl. Acad. Sci. USA *84*, 5216–5220.

Cruz-Vera, L.R., Sachs, M.S., Squires, C.L., and Yanofsky, C. (2011). Nascent polypeptide sequences that influence ribosome function. Curr. Opin. Microbiol. *14*, 160–166.

Del Alamo, M., Hogan, D.J., Pechmann, S., Albanese, V., Brown, P.O., and Frydman, J. (2011). Defining the Specificity of Cotranslationally Acting Chaperones by Systematic Analysis of mRNAs Associated with Ribosome-Nascent Chain Complexes. PLoS Biol. *9*, e1001100.

Deuerling, E., Patzelt, H., Vorderwülbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H., and Bukau, B. (2003). Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. Mol. Microbiol. *47*, 1317–1328.

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.

Frank, J. (2003). Toward an understanding of the structural basis of translation. Genome Biol. *4*, 237.

Gautschi, M., Lilie, H., Fünfschilling, U., Mun, A., Ross, S., Lithgow, T., Rücknagel, P., and Rospert, S. (2001). RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zuotin. Proc. Natl. Acad. Sci. USA *98*, 3762–3767.

Gautschi, M., Mun, A., Ross, S., and Rospert, S. (2002). A functional chaperone triad on the yeast ribosome. Proc. Natl. Acad. Sci. USA *99*, 4209–4214.

Genevaux, P., Keppel, F., Schwager, F., Langendijk-Genevaux, P.S., Hartl, F.U., and Georgopoulos, C. (2004). In vivo analysis of the overlapping functions of DnaK and trigger factor. EMBO Rep. *5*, 195–200.

Gong, F., and Yanofsky, C. (2002). Instruction of translating ribosome by nascent peptide. Science *297*, 1864–1867.

Guthrie, B., and Wickner, W. (1990). Trigger factor depletion or overproduction causes defective cell division but does not block protein export. J. Bacteriol. *172*, 5555–5562.

Hartl, F.U., and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. Nat. Struct. Mol. Biol. *16*, 574–581.

Hesterkamp, T., Hauser, S., Lütcke, H., and Bukau, B. (1996). Escherichia coli trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. Proc. Natl. Acad. Sci. USA *93*, 4437–4441.

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. Biochim. Biophys. Acta *1803*, 650–661.

Hoffmann, A., Merz, F., Rutkowska, A., Zachmann-Brand, B., Deuerling, E., and Bukau, B. (2006). Trigger factor forms a protective shield for nascent polypeptides at the ribosome. J. Biol. Chem. *281*, 6539–6545.

Hundley, H.A., Walter, W., Bairstow, S., and Craig, E.A. (2005). Human Mpp11 J protein: ribosome-tethered molecular chaperones are ubiquitous. Science *308*, 1032–1034.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Kaiser, C.M., Chang, H.-C., Agashe, V.R., Lakshmipathy, S.K., Etchells, S.A., Hayer-Hartl, M., Hartl, F.U., and Barral, J.M. (2006). Real-time observation of trigger factor function on translating ribosomes. Nature *444*, 455–460.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.-W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science *315*, 525–528.

Koplin, A., Preissler, S., Ilina, Y., Koch, M., Scior, A., Erhardt, M., and Deuerling, E. (2010). A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. J. Cell. Biol. *189*, 57–68.

Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat. Struct. Mol. Biol. *16*, 589–597.

Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E., and Bukau, B. (2002). L23 protein functions as a chaperone docking site on the ribosome. Nature *419*, 171–174.

Lakshmipathy, S.K., Tomic, S., Kaiser, C.M., Chang, H.-C., Genevaux, P., Georgopoulos, C., Barral, J.M., Johnson, A.E., Hartl, F.U., and Etchells, S.A. (2007). Identification of nascent chain interaction sites on trigger factor. J. Biol. Chem. *282*, 12186–12193.

Lee, H.C., and Bernstein, H.D. (2001). The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. Proc. Natl. Acad. Sci. USA *98*, 3471–3476.

Lee, H.C., and Bernstein, H.D. (2002). Trigger factor retards protein export in Escherichia coli. J. Biol. Chem. *277*, 43527–43535.

Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Lill, R., Crooke, E., Guthrie, B., and Wickner, W. (1988). The "trigger factor cycle" includes ribosomes, presecretory proteins, and the plasma membrane. Cell *54*, 1013–1018.

Maier, R., Eckert, B., Scholz, C., Lilie, H., and Schmid, F.-X. (2003). Interaction of trigger factor with the ribosome. J. Mol. Biol. *326*, 585–592.

Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. Nature *466*, 1120–1124.

Martinez-Hackert, E., and Hendrickson, W.A. (2009). Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. Cell *138*, 923–934.

Merz, F., Boehringer, D., Schaffitzel, C., Preissler, S., Hoffmann, A., Maier, T., Rutkowska, A., Lozza, J., Ban, N., Bukau, B.*, et al.* (2008). Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. EMBO J. *27*, 1622–1632.

Merz, F., Hoffmann, A., Rutkowska, A., Zachmann-Brand, B., Bukau, B., and Deuerling, E. (2006). The C-terminal domain of Escherichia coli trigger factor represents the central module of its chaperone activity. J. Biol. Chem. *281*, 31963–31971.

Nakatogawa, H., and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. Cell *108*, 629–636.

Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J.C., and Brierley, I. (2006). A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature *441*, 244–247.

Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G.*, et al.* (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. Cell *147*, 1295–1308.

Patzelt, H., Kramer, G., Rauch, T., Schönfeld, H.-J., Bukau, B., and Deuerling, E. (2002). Three-state equilibrium of Escherichia coli trigger factor. Biol. Chem. *383*, 1611–1619.

Patzelt, H., Rüdiger, S., Brehmer, D., Kramer, G., Vorderwülbecke, S., Schaffitzel, E., Waitz, A., Hesterkamp, T., Dong, L., Schneider-Mergener, J.*, et al.* (2001). Binding specificity of Escherichia coli trigger factor. Proc. Natl. Acad. Sci. USA *98*, 14244–14249.

Peisker, K., Chiabudini, M., and Rospert, S. (2010). The ribosome-bound Hsp70 homolog Ssb of Saccharomyces cerevisiae. Biochim. Biophys. Acta *1803*, 662–672.

Preissler, S., and Deuerling, E. (2012). Ribosome-associated chaperones as key players in proteostasis. Trends Biochem. Sci. *37*, 274–283.

Raine, A., Lovmar, M., Wikberg, J., and Ehrenberg, M. (2006). Trigger factor binding to ribosomes with nascent peptide chains of varying lengths and sequences. J. Biol. Chem. *281*, 28033–28038.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. Cell *108*, 557–572.

Rauch, T., Hundley, H.A., Pfund, C., Wegrzyn, R.D., Walter, W., Kramer, G., Kim, S.-Y., Craig, E.A., and Deuerling, E. (2005). Dissecting functional similarities of ribosome-associated chaperones from Saccharomyces cerevisiae and Escherichia coli. Mol. Microbiol. *57*, 357–365.

Rodnina, M.V., Beringer, M., and Wintermeyer, W. (2007). How ribosomes make peptide bonds. Trends Biochem. Sci. *32*, 20–26.

Rospert, S., Dubaquié, Y., and Gautschi, M. (2002). Nascent-polypeptide-associated complex. Cell. Mol. Life Sci. *59*, 1632–1639.

Rüdiger, 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.

Rutkowska, A., Mayer, M.P., Hoffmann, A., Merz, F., Zachmann-Brand, B., Schaffitzel, C., Ban, N., Deuerling, E., and Bukau, B. (2008). Dynamics of trigger factor interaction with translating ribosomes. J. Biol. Chem. *283*, 4124–4132.

Sakr, S., Cirinesi, A.-M., Ullers, R.S., Schwager, F., Georgopoulos, C., and Genevaux, P. (2010). Lon protease quality control of presecretory proteins in Escherichia coli and its dependence on the SecB and DnaJ (Hsp40) chaperones. J. Biol. Chem. *285*, 23506–23514.

Scholz, C., Stoller, G., Zarnt, T., Fischer, G., and Schmid, F.X. (1997). Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. EMBO J. *16*, 54–58.

Sørensen, M.A., Kurland, C.G., and Pedersen, S. (1989). Codon usage determines translation rate in Escherichia coli. J. Mol. Biol. *207*, 365–377.

Stoller, G., Rücknagel, K.P., Nierhaus, K.H., Schmid, F.X., Fischer, G., and Rahfeld, J.U. (1995). A ribosome-associated peptidyl-prolyl cis/trans isomerase identified as the trigger factor. EMBO J. *14*, 4939–4948.

Teter, S.A., Houry, W.A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C., and Hartl, F.U. (1999). Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell *97*, 755–765.

Tomic, S., Johnson, A.E., Hartl, F.U., and Etchells, S.A. (2006). Exploring the capacity of trigger factor to function as a shield for ribosome bound polypeptide chains. FEBS Lett. *580*, 72–76.

Ullers, R.S., Ang, D., Schwager, F., Georgopoulos, C., and Genevaux, P. (2007). Trigger Factor can antagonize both SecB and DnaK/DnaJ chaperone functions in Escherichia coli. Proc. Natl. Acad. Sci. USA *104*, 3101–3106.

Ullers, R.S., Luirink, J., Harms, N., Schwager, F., Georgopoulos, C., and Genevaux, P. (2004). SecB is a bona fide generalized chaperone in Escherichia coli. Proc. Natl. Acad. Sci. USA *101*, 7583–7588.

Valent, Q.A., Kendall, D.A., High, S., Kusters, R., Oudega, B., and Luirink, J. (1995). Early events in preprotein recognition in E. coli: interaction of SRP and trigger factor with nascent polypeptides. EMBO J. *14*, 5494–5505.

Vorderwülbecke, S., Kramer, G., Merz, F., Kurz, T.A., Rauch, T., Zachmann-Brand, B., Bukau, B., and Deuerling, E. (2004). Low temperature or GroEL/ES overproduction permits growth of Escherichia coli cells lacking trigger factor and DnaK. FEBS Lett. *559*, 181–187.

Yanagitani, K., Kimata, Y., Kadokura, H., and Kohno, K. (2011). Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. Science *331*, 586–589.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol. *16*, 274–280.

Zimmerman, S.B., and Trach, S.O. (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. J. Mol. Biol. *222*, 599–620.

CHAPTER TWO

Review on the Biology of Nascent Chains

Introduction

Ribosomes are molecular machines that accelerate the formation of peptides bonds. In exponentially growing cells, most ribosomes are actively engaged in translation, so the rate of protein production effectively limits how fast cells can grow. Indeed, bacteria possess ribosomes that catalyze, on average, twenty amino acids per second (Frank, 2003), and accordingly divide more rapidly than eukaryotes whose ribosomes are intrinsically slower, catalyzing roughly five amino acids per second (Thompson and Heywood, 1974). Yet local fluctuations in translation speed can have profound consequences, with transient pausing dictating numerous co-translational processes, including controlling the fold (Kimchi-Sarfaty et al., 2007; Zhang et al., 2009) and localization (Mariappan et al., 2010) of the protein being synthesized. So it is unsurprising that numerous control mechanisms exist throughout this fundamental process.

Compelling evidence now shows that the nascent chain has the capacity to alter its own fate (Kramer et al., 2009), suggesting that the ribosome itself can be modulated by the emerging polypeptide. Selective regulation of the ribosome by the nascent chain, as well as the use of specialized ribosomes (i.e. those that have distinct ribosomal proteins and modifications) (Xue and Barna, 2012), belong to a new arsenal of mechanisms the ribosome can exploit to express the diversity of products while using the same overall machinery. By and large, this review focuses on the various principles governing nascent chain-mediated regulation of the ribosome.

Nascent Chain Interactions with the Exit Tunnel Drive Physiological Processes

Nascent chains are driven through the exit tunnel by newly incorporated amino acids. The tunnel houses critical structural features that affect various aspects of

translation control (Kramer et al., 2009), and in no way acts as an inert conduit as once believed (Ban et al., 2000). In bacteria, this tunnel measures ~100 Å in length (Ban et al., 2000), being able to contain a polypeptide composed of roughly 30 residues when fully extended in conformation. Although decorated with segments of L4, L22, and L23, the tunnel wall is predominately lined by the 23S rRNA. Hence, its electrostatic potential is overwhelmingly negative (Lu et al., 2007), suggesting that the tunnel itself is capable of modulating local interactions with positively charged residues (Lu and Deutsch, 2008). Indeed, expression of polybasic chains force elongation arrest in eukaryotes (Dimitrova et al., 2009), activating the proteasome-mediated destruction of the polypeptide along with degradation of the template by exonucleases. This quality control mechanism called non-stop decay is triggered when transcripts lack stop codons (Akimitsu, 2008), translating polylysines encoded by the polyadenylated tail as a result.

Remarkably, nascent chain-induced translation arrest can even regulate protein production. In one such example, stalling at *secM* controls the expression of the translocation ATPase SecA in response to the secretion status of the cell (Murakami et al., 2004). Here, *secM* and *secA* are fully transcribed as an operon, but hairpin formation overlapping the ribosome recruitment site responsible for SecA expression lowers its own frequency of initiation (Nakatogawa et al., 2004). Since lack of SecA prevents SecM export (Nakatogawa et al., 2005), prolonged stalling at *secM* simply increases the time at which this ribosome recruitment site is accessible, ultimately raising SecA production. Operating under negative feedback, release of elongation arrest by SecM occurs upon the pulling force generated by newly expressed SecA molecules (Butkus et al., 2003). In a second example, stalling at *tnaC*, mediated by excess levels of tryptophan (Cruz-Vera et al., 2006), occludes recognition of the rho-dependent termination site (Gong and Yanofsky, 2002), enabling the complete transcription of the polycistronic message that

also encodes for tryptophanase and tryptophan permease. Release of TnaC-induced stalling occurs only when tryptophan is catabolized and/or exported by these newly expressed enzymes (Cruz-Vera et al., 2005).

Yet how do these nascent chains mechanistically inhibit the ribosome? To address this matter, structure models were derived from cryo-EM reconstructions of both SecM- and TnaC-arrested ribosome complexes (Bhushan et al., 2011; Seidelt et al., 2009). Both nascent chains made numerous contacts with residues of the tunnel wall. Yet neither complex showed considerable rearrangement in tunnel structure, implying that the ribosome itself does not relay the cues necessary for inactivating peptidyl transferase activity. Previous mutation analyses established that interactions between critical residues of either nascent chain and the constriction aperture were required for stalling (Cruz-Vera et al., 2005; Nakatogawa and Ito, 2002). So interactions at this junction, located ~30 Å away from the catalytic center, ought to lower the conformational entropy (or sampling of folds) of the nascent chain, specifying a more rigid structure in which the nascent chain could act to shut off catalytic activity. Indeed, local perturbations in the active site were observed for both arrested complexes, likely mediated by the constrained nascent chain. In the SecM-arrested complex (Bhushan et al., 2011), the position of the peptidyl-linked tRNA was shifted by ~2 Å, preventing the nucleophilic attack from the A-site prolyl tRNA. Yet surprisingly, this proline residue is absolutely required for arrest (Muto et al., 2006; Nakatogawa and Ito, 2002). This constraint is explained by the lower nucleophilicity inherent to the N-alkyl group of prolines (Pavlov et al., 2009) when compared with all other amino acids. In the TnaC-stalled complex by contrast (Seidelt et al., 2009), two conserved residues present in the peptidyl transferase center (A2602 and U2585) adopted conformations inhibiting the association of release factor.

Even before emerging from the tunnel, nascent chains induce recruitment of the signal recognition particle (SRP) to the ribosome-nascent chain complex (RNC) at L23 in bacteria (Gu et al., 2003) or L25 in eukaryotes. More precisely, the translation of short chains (composed of ~30 residues) increases the affinity of SRP association by ~100-fold when compared with binding to vacant ribosomes (Bornemann et al., 2008). Strikingly, short chains with deficient signal sequences induce SRP recruitment just as well as those with functional ones. By contrast, long chains with accessible signal sequences stably associate SRP molecules to RNCs only if they are export competent. So how do nascent chains signal beyond the tunnel for SRP recruitment? Mutation analyses suggest that L23 itself modulates this allostery (Bornemann et al., 2008). Although located near the site of peptide exit, L23 possesses a flexible loop that penetrates the tunnel wall ~60–70 Å from the peptidyl transferase center. So it is likely that this loop senses the incoming nascent chain, ultimately increasing affinity of SRP through the change in conformation at its cytoplasmic face. Indeed, RNCs with intraloop deletions of L23 decrease SRP recruitment only when carrying short chains (Bornemann et al., 2008).

Co-translational Folding of the Nascent Chain

Nascent chains can form alpha helices even before they exit the ribosome. In order to observe secondary structure formation inside the tunnel, studies relied on mass tagging the nascent chain through pegylation of the only accessible cysteine (Lu and Deutsch, 2005). By inserting polyalanine stretches along the nascent chain, this polypeptide would effectively serve as a molecular tape measure. Surprisingly, only those nascent chains with polyalanines most proximal to the peptidyl transferase center saw reduced levels of pegylation, suggesting that these alpha helices unfold before

exiting. It could be that this structure plays an important role in signaling, but this hypothesis requires further validation. More recently, alpha helical nascent chains were directly visualized by cryo-EM reconstruction of translating RNCs (Bhushan et al., 2010). Here, alpha helix formation was observed at the very end of the tunnel, suggesting that nascent chains can initiate the folding process even before exiting the ribosome. Pre-folding of the nascent chain likely reduces the number of conformations the polypeptide can sample, ultimately guiding the correct fold of the overall protein.

Interplay of Ribosome Associated-Factors At the Nascent Chain Exit Platform

As soon as nascent chains emerge from the tunnel, they are greeted by an assortment of ribosome-associated factors (RAFs) (Kramer et al., 2009), including processing enzymes, targeting factors, and molecular chaperones. This remarkable orchestration of RAF binding, regulated both spatially and in time, takes place on the tunnel platform, which surrounds the site of peptide exit. This platform features three universally conserved proteins: L23, L24, and L29 (designated as L25, L26, and L35 in eukaryotes). Surprisingly, L23 (L25) is responsible for the majority of factor recruitment, interacting with SecA (Huber et al., 2011) and trigger factor (TF) (Kramer et al., 2002) in bacteria, as well as nascent chain-associated complex (NAC) (Wegrzyn et al., 2006) and N-acetyltransferase (Polevoda et al., 2008) in eukaryotes. Moreover, L23 (L25) can recruit both bacterial and eukaryotic homologues of SRP (Gu et al., 2003; Halic et al., 2004) and the translocon (Becker et al., 2009; Mitra et al., 2005).

How do these factors synchronize their activities around this crowded platform? Does steric occlusion force competitive association or can the platform freely accommodate each RAF? Structure models predict steric clashing between SRP and TF at L23 (Merz et al., 2008; Schaffitzel et al., 2006). Yet both can associate with RNCs at

the same time in vitro (Buskiewicz et al., 2004; Raine et al., 2004). Thus considerable changes in structure must occur for their simultaneous accommodation. Indeed, the conformation of SRP changes when interacting with RNCs (Buskiewicz et al., 2009). N-terminal signal sequences strongly stabilize SRP binding, even when TF is supplied in excess. Yet crosslinking experiments showed that SRP and TF could compete for signal anchor sequences (i.e. those that are not cleaved following translocation) (Beck et al., 2000; Eisner et al., 2006; Ullers et al., 2003; Valent et al., 1995). Remarkably, TF could fully displace SRP-bound RNCs harboring these nascent chains. The significance of these in vitro findings are bolstered by in vivo observations, showing that lack of TF accelerates protein export (Lee and Bernstein, 2002), bypasses the requirement for presecretory targeting chaperone SecB (Ullers et al., 2007), and increases the fraction of membrane-bound ribosomes (Ullers et al., 2007).

In bacteria, peptidyl deformylases (PDFs) excise the formyl groups of N-terminal methionines, but this enzymatic step must occur before methionine aminopeptidases (MAPs) can act (Giglione et al., 2004). Interestingly, structure models show that TF and PDF do not sterically compete for the same binding space (Bingel-Erlenmeyer et al., 2008). PDF inserts its C-terminal extension into a crevice between L22 and L32, with its active site rotated toward the tunnel exit. Consequently, RNC association is essential for PDF processing activity (Bingel-Erlenmeyer et al., 2008). Currently, it is unknown where MAP lies in relation to PDF and TF. So far, in vitro studies have shown that MAPs can process RNCs harboring ~40 residue long nascent chains (Ball and Kaesberg, 1973).

Eukaryotes possess two ribosome-associated chaperone systems: Ssb/Ssz/Zuotin (Gautschi et al., 2001) and NAC (Rospert et al., 2002). Even though Zuotin crosslinks to L31 (Gautschi et al., 2001), the precise location of Ssb/Ssz/Zuotin binding remains unknown. Ssb is an Hsp70 homologue possessing ATPase activity that

is further stimulated by Ssz/Zuotin (Huang et al., 2005). By contrast, mechanisms that govern chaperone activity of NAC are presently unknown. Nonetheless, deletion of both NAC and Ssb is synthetic under conditions that exacerbate folding stress (Koplin et al., 2010). Intriguingly, the absence of NAC enables the binding of SRP with RNCs lacking signal sequences in vitro (Lauring et al., 1995), suggesting that NAC might modulate the targeting fidelity by increasing SRP specificity. Yet how these partners cooperate at the platform remains an open question.

Perspective

Nascent chains enforce the selective regulation of ribosomes by halting erroneously translated messages, driving gene expression through translation arrest, and inducing factor recruitment. Yet it remains to be seen if other processes are modulated by interactions between the nascent chain and tunnel wall. Furthermore, future work will likely address how RAFs coordinate their actions to ensure the production of functional polypeptides. So far, it is clear that the interplay between these RAFs is dynamic and complex. But full understanding of the regulatory principles that govern nascent chain biology is far from complete.

REFERENCES

Akimitsu, N. (2008). Messenger RNA surveillance systems monitoring proper translation termination. J. Biochem. *143*, 1–8.

Ball, L.A., and Kaesberg, P. (1973). Cleavage of the N-terminal formylmethionine residue from a bacteriophage coat protein in vitro. J. Mol. Biol. *79*, 531–537.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science *289*, 905–920.

Beck, K., Wu, L.F., Brunner, J., and Müller, M. (2000). Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. EMBO J. *19*, 134–143.

Becker, T., Bhushan, S., Jarasch, A., Armache, J.-P., Funes, S., Jossinet, F., Gumbart, J., Mielke, T., Berninghausen, O., Schulten, K., et al. (2009). Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. Science *326*, 1369–1373.

Bhushan, S., Gartmann, M., Halic, M., Armache, J.-P., Jarasch, A., Mielke, T., Berninghausen, O., Wilson, D.N., and Beckmann, R. (2010). alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. Nat. Struct. Mol. Biol. *17*, 313–317.

Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., Wilson, D.N., and Beckmann, R. (2011). SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. PLoS Biol. *9*, e1000581.

Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolić, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B., and Ban, N. (2008). A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature *452*, 108–111.

Bornemann, T., Jöckel, J., Rodnina, M.V., and Wintermeyer, W. (2008). Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. Nat. Struct. Mol. Biol. *15*, 494–499.

Buskiewicz, I., Deuerling, E., Gu, S.-Q., Jöckel, J., Rodnina, M.V., Bukau, B., and Wintermeyer, W. (2004). Trigger factor binds to ribosome-signal-recognition particle (SRP) complexes and is excluded by binding of the SRP receptor. Proc. Natl. Acad. Sci. USA *101*, 7902–7906.

Buskiewicz, I.A., Jöckel, J., Rodnina, M.V., and Wintermeyer, W. (2009). Conformation of the signal recognition particle in ribosomal targeting complexes. RNA *15*, 44–54.

Butkus, M.E., Prundeanu, L.B., and Oliver, D.B. (2003). Translocon "pulling" of nascent SecM controls the duration of its translational pause and secretion-responsive secA regulation. J. Bacteriol. *185*, 6719–6722.

Cruz-Vera, L.R., Gong, M., and Yanofsky, C. (2006). Changes produced by bound tryptophan in the ribosome peptidyl transferase center in response to TnaC, a nascent leader peptide. Proc. Natl. Acad. Sci. USA *103*, 3598–3603.

Cruz-Vera, L.R., Rajagopal, S., Squires, C., and Yanofsky, C. (2005). Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression. Mol. Cell *19*, 333–343.

Dimitrova, L.N., Kuroha, K., Tatematsu, T., and Inada, T. (2009). Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. J. Biol. Chem. *284*, 10343–10352.

Eisner, G., Moser, M., Schäfer, U., Beck, K., and Müller, M. (2006). Alternate recruitment of signal recognition particle and trigger factor to the signal sequence of a growing nascent polypeptide. J. Biol. Chem. *281*, 7172–7179.

Frank, J. (2003). Toward an understanding of the structural basis of translation. Genome Biol. *4*, 237.

Gautschi, M., Lilie, H., Fünfschilling, U., Mun, A., Ross, S., Lithgow, T., Rücknagel, P., and Rospert, S. (2001). RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zuotin. Proc. Natl. Acad. Sci. USA *98*, 3762–3767.

Giglione, C., Boularot, A., and Meinnel, T. (2004). Protein N-terminal methionine excision. Cell. Mol. Life Sci. *61*, 1455–1474.

Gong, F., and Yanofsky, C. (2002). Instruction of translating ribosome by nascent peptide. Science *297*, 1864–1867.

Gu, S.-Q., Peske, F., Wieden, H.-J., Rodnina, M.V., and Wintermeyer, W. (2003). The signal recognition particle binds to protein L23 at the peptide exit of the Escherichia coli ribosome. RNA *9*, 566–573.

Halic, M., Becker, T., Pool, M.R., Spahn, C.M., Grassucci, R.A., Frank, J., and Beckmann, R. (2004) Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature *427*, 808–814.

Huang, P., Gautschi, M., Walter, W., Rospert, S., and Craig, E.A. (2005). The Hsp70 Ssz1 modulates the function of the ribosome-associated J-protein Zuo1. Nat. Struct. Mol. Biol. *12*, 497–504.

Huber, D., Rajagopalan, N., Preissler, S., Rocco, M.A., Merz, F., Kramer, G., and Bukau, B. (2011). SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria. Mol. Cell *41*, 343–353.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.-W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science *315*, 525–528.

Koplin, A., Preissler, S., Ilina, Y., Koch, M., Scior, A., Erhardt, M., and Deuerling, E. (2010). A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. J. Cell Biol. *189*, 57–68.

Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat. Struct. Mol. Biol. *16*, 589–597.

Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E., and Bukau, B. (2002). L23 protein functions as a chaperone docking site on the ribosome. Nature *419*, 171–174.

Lauring, B., Sakai, H., Kreibich, G., and Wiedmann, M. (1995). Nascent polypeptide-associated complex protein prevents mistargeting of nascent chains to the endoplasmic reticulum. Proc. Natl. Acad. Sci. USA *92*, 5411–5415.

Lee, H.C., and Bernstein, H.D. (2002). Trigger factor retards protein export in Escherichia coli. J. Biol. Chem. *277*, 43527–43535.

Lu, J., and Deutsch, C. (2005). Folding zones inside the ribosomal exit tunnel. Nat. Struct. Mol. Biol. *12*, 1123–1129.

Lu, J., and Deutsch, C. (2008). Electrostatics in the ribosomal tunnel modulate chain elongation rates. J. Mol. Biol. *384*, 73–86.

Lu, J., Kobertz, W.R., and Deutsch, C. (2007). Mapping the electrostatic potential within the ribosomal exit tunnel. J. Mol. Biol. *371*, 1378–1391.

Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. Nature *466*, 1120–1124.

Merz, F., Boehringer, D., Schaffitzel, C., Preissler, S., Hoffmann, A., Maier, T., Rutkowska, A., Lozza, J., Ban, N., Bukau, B., et al. (2008). Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. EMBO J. *27*, 1622–1632.

Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C.L., Ban, N., and Frank, J. Structure of the E. coli protein-conducting channel bound to a translating ribosome. Nature *438*, 318–324.

Murakami, A., Nakatogawa, H., and Ito, K. (2004). Translation arrest of SecM is essential for the basal and regulated expression of SecA. Proc. Natl. Acad. Sci. USA *101*, 12330–12335.

Muto, H., Nakatogawa, H., and Ito, K. (2006). Genetically encoded but nonpolypeptide prolyl-tRNA functions in the A site for SecM-mediated ribosomal stall. Mol. Cell *22*, 545–552.

Nakatogawa, H., and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. Cell *108*, 629–636.

Nakatogawa, H., Murakami, A., and Ito, K. (2004). Control of SecA and SecM translation by protein secretion. Curr. Opin. Microbiol. *7*, 145–150.

Nakatogawa, H., Murakami, A., Mori, H., and Ito, K. (2005). SecM facilitates translocase function of SecA by localizing its biosynthesis. Genes Dev. *19*, 436–444.

Pavlov, M.Y., Watts, R.E., Tan, Z., Cornish, V.W., Ehrenberg, M., and Forster, A.C. (2009). Slow peptide bond formation by proline and other N-alkylamino acids in translation. Proc. Natl. Acad. Sci. USA *106*, 50–54.

Polevoda, B., Brown, S., Cardillo, T.S., Rigby, S., and Sherman, F. (2008). Yeast N(alpha)-terminal acetyltransferases are associated with ribosomes. J. Cell. Biochem. *103*, 492–508.

Raine, A., Ivanova, N., Wikberg, J.E.S., and Ehrenberg, M. (2004). Simultaneous binding of trigger factor and signal recognition particle to the E. coli ribosome. Biochimie *86*, 495–500.

Rospert, S., Dubaquié, Y., and Gautschi, M. (2002). Nascent-polypeptide-associated complex. Cell. Mol. Life Sci. *59*, 1632–1639.

Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J.P., Koerten, H.K., Koning, R.I., and Ban, N. (2006). Structure of the E. coli signal recognition particle bound to a translating ribosome. Nature *444*, 503–506.

Seidelt, B., Innis, C.A., Wilson, D.N., Gartmann, M., Armache, J.-P., Villa, E., Trabuco, L.G., Becker, T., Mielke, T., Schulten, K., et al. (2009). Structural insight into nascent polypeptide chain-mediated translational stalling. Science *326*, 1412–1415.

Thompson, W.C., and Heywood, S.M. (1974). Basic aspects of protein synthesis in muscle. J. Anim. Sci. *38*, 1050–1053.

Ullers, R.S., Ang, D., Schwager, F., Georgopoulos, C., and Genevaux, P. (2007). Trigger Factor can antagonize both SecB and DnaK/DnaJ chaperone functions in Escherichia coli. Proc. Natl. Acad. Sci. USA *104*, 3101–3106.

Ullers, R.S., Houben, E.N.G., Raine, A., ten Hagen-Jongman, C.M., Ehrenberg, M., Brunner, J., Oudega, B., Harms, N., and Luirink, J. (2003). Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the Escherichia coli ribosome. J. Cell Biol. *161*, 679–684.

Valent, Q.A., Kendall, D.A., High, S., Kusters, R., Oudega, B., and Luirink, J. (1995). Early events in preprotein recognition in E. coli: interaction of SRP and trigger factor with nascent polypeptides. EMBO J. *14*, 5494–5505.

Wegrzyn, R.D., Hofmann, D., Merz, F., Nikolay, R., Rauch, T., Graf, C., and Deuerling, E. (2006). A conserved motif is prerequisite for the interaction of NAC with ribosomal protein L23 and nascent chains. J. Biol. Chem. *281*, 2847–2857.

Xue, S., and Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat. Rev. Mol. Cell Biol. *13*, 355–369.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol. *16*, 274–280.

CHAPTER THREE

Investigating the Determinants of Ribosome Pausing in Bacterial Translation

The anti-Shine-Dalgarno Sequence Drives Translational Pausing and Codon Choice in Bacteria

Gene-Wei Li,[1] Eugene Oh,[1] and Jonathan S. Weissman[1]

[1] Department of Cellular and Molecular Pharmacology, Howard Hughed Medical Institute, University of California, San Francisco, San Francisco, CA 94158, USA

Reproduced from Nature, *484*(7395), 538–541, Copyright 2012

Introduction

Protein synthesis by ribosomes takes place on a linear substrate but at non-uniform speeds. Transient pausing of ribosomes can affect a variety of cotranslational processes, including protein targeting and folding (Kramer et al., 2009). These pauses are influenced by the sequence of the messenger RNA (Plotkin and Kudla, 2011). Thus, redundancy in the genetic code allows the same protein to be translated at different rates. However, our knowledge of both the position and the mechanism of translational pausing in vivo is highly limited. Here, we present a genome-wide analysis of translational pausing in bacteria by ribosome profiling—deep sequencing of ribosome-protected mRNA fragments (Ingolia et al., 2009; Ingolia et al., 2011; Oh et al., 2011). This approach enables the high resolution measurement of ribosome density profiles along most transcripts at unperturbed, endogenous expression levels. Unexpectedly, we found that codons decoded by rare transfer RNAs do not lead to slow translation under nutrient rich conditions. Instead, Shine-Dalgarno-(SD)-(Shine and Dalgarno, 1974)-like features within coding sequences cause pervasive translational pausing. Using an orthogonal ribosome (Hui and de Boer, 1987; Rackham and Chin, 2005) possessing an altered anti-SD sequence, we show that pausing is due to hybridization between the mRNA and 16S ribosomal RNA of the translating ribosome. In protein coding sequences, internal SD sequences are disfavored, which leads to biased usage, avoiding codons and codon pairs that resemble canonical SD sites. Our results indicate that internal SD-like sequences are a major determinant of translation rates and a global driving force for the coding of bacterial genomes.

Results and Discussion

Our current understanding of sequence dependent translation rates in vivo

derives largely from pioneering work begun in the 1980s (Andersson and Kurland, 1990; Pedersen, 1984; Sorensen et al., 1989; Sorensen and Pedersen, 1991; Varenne et al., 1984). These studies, which measured protein synthesis times using pulse labeling, established that different mRNAs could be translated with different elongation rates. In particular, messages decoded by less abundant tRNAs were translated slowly, although this effect was exaggerated by the overexpression of mRNA, which can lead to the depletion of available tRNAs (Pedersen, 1984). Even with fixed tRNA usage, different synonymously coded mRNAs were translated at different rates (Sorensen and Pedersen, 1991). This result, together with the observation of biased occurrence of adjacent codon pairs (Gutman and Hatfield, 1989), suggested that tRNA abundance is not the only determinant of elongation rates. Further investigations addressing what determines the rate of translation in vivo, however, have been hampered by the limited temporal and positional resolution of existing techniques.

To provide a high resolution view of local translation rates, we used the recently developed ribosome profiling strategy (Ingolia et al., 2009; Ingolia et al., 2011; Oh et al., 2011) to map ribosome occupancy along each mRNA (Figures S1). We focused on two distantly related bacterial species, the Gram-negative bacterium *Escherichia coli* and the Gram-positive bacterium *Bacillus subtilis*. To preserve the state of translation, cells were flash frozen in liquid nitrogen after the rapid filtration of exponential phase cultures. Ribosome-protected footprints were generated through nuclease treatment of cell extract in the presence of inhibitors of translation elongation (see Extended Experimental Procedures). These steps ensured that most ribosomes were polysome-associated after lysis and stayed assembled as 70S particles during digestion (Figure S2). After deep sequencing, 2,257 genes from *E. coli* and 1,580 genes from *B. subtilis* had an average coverage of at least ten sequencing reads per codon. The observed variability of

ribosome footprint profiles across individual genes was highly reproducible (r = 0.99 between biological replicates) (Figure S3).

Several observations argued that ribosome transit time is proportional to the occupancy at each position. First, we observed negligible internal initiation and early termination associated with ribosome pause sites (Figure S4). Second, ribosomes remained intact during footprinting, which enabled the large majority of ribosome-protected fragments to be captured (Figure S2). Third, the variability introduced during the conversion of RNA fragments into a sequenceable DNA library contributed minimally to our measures of variability in ribosome occupancy (Figure S5).

With our genome-wide view of local translation rates, we confirmed established examples of peptide-mediated stalling at transcripts *secM* (Nakatogawa and Ito, 2002) and *tnaC* (Gong and Yanofsky, 2002) in *E. coli* and *mifM* (Chiba et al., 2011) in *B. subtilis* (Figures 1A and S6). Strikingly, in addition to these known stalling sites, the observed ribosome occupancy was highly variable across coding regions, as illustrated for *secA* (Figure 1A). We found that ribosome density often surpasses by more than tenfold the mean density, and the vast majority of these translational pauses are uncharacterized.

We first sought to determine whether the identity of the codon being decoded could account for the differences in local translation rates, by examining the average ribosome occupancy for each of the 61 codons in the ribosomal A-site. Surprisingly, there was little correlation between the average occupancy of a codon and existing abundance measurements of the corresponding tRNAs (Dong et al., 1996) (Figures 1B, 1C, and S7). Most notably, the six serine codons had the highest ribosome occupancy for *E. coli* cultured in Luria broth (Figure 1B). Because serine is the first amino acid to be catabolized by *E. coli* when sugar is absent (Pruss et al., 1994; Sezonov et al., 2007),

we reasoned that the increased ribosome occupancy might be due to limited serine supply. Indeed, serine-associated pauses were greatly decreased in glucose-supplemented MOPS medium (Figure 1C). The increase in serine codon occupancy when glucose becomes limiting confirmed our ability to capture translation rates at each codon. However, the identity of the A-site codon, which had less than a twofold effect on ribosome occupancy (Figure 1C), could not account for the large variability in ribosome density along messages.

What, then, are the sequence features that cause slow translation? Without a priori knowledge about where such features would be located relative to the ribosomal A-site, we calculated the cross-correlation function between intragenic ribosome occupancy profiles and the presence of a given trinucleotide sequence on the mRNA independent of reading frame. Strong correlation was observed for six trinucleotide sequences (Figure 1D) that resembled features found in Shine-Dalgarno (SD) sequences (Shine and Dalgarno, 1974). The highest correlation occurred when the SD-like feature was 8–11 nucleotides upstream of the position occupied by the ribosomal A-site. This spacing coincides with the optimal spacing for ribosome binding at start codons (Chen et al., 1994). However, unlike canonical SD sites, which enable initiation of translation, the observed pauses were associated with SD-like features within the body of coding regions. The accumulation of ribosomes at internal SD-like sequences was observed across two divergent phyla of bacteria (Figure 2A), suggesting that the phenomenon occurs generally in bacterial species. The same correlation was not observed for the budding yeast *Saccharomyces cerevisiae*, whose ribosomes, like those of other eukaryotes, do not contain an anti-SD (aSD) site.

As predicted by a model in which the interaction between mRNA and the aSD site of the 16S rRNA drives pausing, the predicted hybridization free energy of a

hexanucleotide to the aSD sequence was a strong indicator of its average downstream ribosome occupancy (Figure 2B). Furthermore, there was a clear correspondence on individual transcripts between SD-like sequences and pauses. For example, Figure 2C shows that in *ompF*, individual SD-like features are associated with elevated ribosome occupancy 8–11 nucleotides downstream. Moreover, a synonymous mutation (GGUGGU to GGCGGC) that decreased the affinity for the aSD site led to reduced ribosome occupancy specifically at the mutated sequence, suggesting a causal relationship between the SD-like feature and the excess ribosome density.

We next sought to evaluate directly whether the excess footprint density seen at internal SD-like sequences was due to pausing of elongating ribosomes rather than attempted internal initiation, driven by SD–aSD interactions (Figure 3A). To distinguish between these possibilities, we used a previously described orthogonal ribosome (O-ribosome) system in which a mutant form of the 16S rRNA with an altered aSD site is expressed together with wild-type 16S rRNA (Rackham and Chin, 2005). O-ribosomes containing the mutant 16S RNA will only translate a target mRNA that has the corresponding orthogonal SD (O-SD) sequence before its start codon. Conversely, a message whose translation is driven by the O-SD sequence will only be translated by O-ribosomes, and not by wild-type ribosomes. This system thus allows one to determine the source of regions of excess ribosome footprints, because elongating O-ribosomes would pause at internal O-SD sequences, whereas attempted internal initiation would still occur at SD sequences as a result of the cellular pool of wild-type ribosomes.

We compared the ribosome occupancy profiles of a *lacZ* message that was translated by either O-ribosomes or wild-type ribosomes. The occupancy profile of the *lacZ* message exclusively translated by O-ribosomes was correlated with O-SD-like features, and not with SD-like features (Figure 3C). This is in marked contrast with the

same *lacZ* sequence translated by wild-type ribosomes (Figure 3B). As an internal control in O-ribosome expressing cells, all other genes, which were translated by wild-type ribosomes, still maintained SD-correlated ribosome occupancy profiles (Figure 3C). These observations established that elongating ribosomes pause during protein synthesis and that hybridization between the aSD site in the elongating ribosome and internal SD-like sequences gives rise to these pauses.

Global analysis of pause sites revealed that internal SD-like sequences are the dominant feature controlling translational pausing: about 70% of the strong pauses (that is, those that have ribosome occupancies more than tenfold over the mean) are associated with SD sites (Figure S8). Although the interaction between internal SD sequences in a message and elongating ribosomes has been documented in specialized cases, including promoting frameshifting in vivo (Larsen et al., 1994; Weiss et al., 1988) and ribosome stalling in single molecule experiments in vitro (Wen et al., 2008), there was little indication that internal SD-like sequences are the major determinant of elongation rate during translation.

Because translational pausing limits the amount of free ribosomes available for initiating protein synthesis, widespread internal SD-like sequences could decrease bacterial growth rates. Accordingly, we found that strong SD-like sequences are generally avoided in the coding region of *E. coli* genes: hexamer sequences that strongly bind aSD sites are universally rare, whereas low affinity hexamers have variable rates of occurrence (Figure 4A). Consistent with translational pausing being the driving force for this bias, depletion of SD-like sequences was observed only in protein coding genes, and not in genes encoding rRNA or tRNA (Figure S9). The selection against SD-like features in turn impacts both synonymous codon choice and codon pair choice. At the codon level, SD-like codons GAG, AGG and GGG are all minor codons in *E.* coli and *B.*

subtilis. The evolutionary origin of codon selection is often attributed to differences in tRNA abundance (Ikemura, 1981; Plotkin and Kudla, 2011) because its level is correlated with codon usage (Dong et al., 1996). Instead, we propose that SD-like codons are disfavored as a result of their interactions with rRNA, and that tRNA expression levels followed codon adaptation.

At the codon pair level, we can now account for the selection against two consecutive codons that resemble SD sequences. This is illustrated for Gly-Gly pairs, which are coded by GGNGGN sequences (Figure 4B). The most abundant Gly-Gly coding sequence, GGCGGC, has the lowest affinity for the aSD sequence, whereas Gly-Gly coding sequences that strongly resemble SD sites, including GGAGGU, which perfectly complements the aSD site, rarely appear. This underrepresentation holds even after correcting for the usage of individual codons (Figure 4B); for example, GGAGGU is considerably less common than GGUGGA. Other amino acid pairs that can be coded with strong SD sites also showed the same bias (Figure S10). The preference in codon pairs stems from the sequence identity and not codon identity, because the same trend is seen in hexamers that are not aligned to codon pairs (Figure S11). Although not every bias in codon pair usage can be explained here, the disadvantage associated with SD-induced translational pausing offers a clear mechanistic view of why certain codon pairs are avoided.

Despite the selection against internal SD-like sequences, they remain a major driving force for translational pausing. In addition, we found similar pausing patterns between conserved genes in *E. coli* and *B. subtilis* (Figure 4C). For an mRNA encoding a specific protein, it may not be possible to fully eliminate sequences with affinity for the aSD site without changing the peptide sequence. For example, in the case of Gly-Gly, even the GGCGGC pair has substantial affinity for the aSD site. The optimization for

translation rate therefore cannot be achieved only at the level of mRNA coding: it is also constrained by the requirement to make a functional peptide sequence.

The observation that the ability of elongating ribosomes to interact with SD-like sequences is highly conserved suggests that this mechanism of pausing is exploited for functional purposes. Indeed, a highly conserved internal SD site exists in the gene encoding peptide chain release factor 2 (RF2) (Baranov et al., 2002). This sequence has an important function in promoting a translational frameshift to enable its expression. In addition, pausing at internal SD-like sites could modulate the cotranslational folding of the nascent peptide chain (Figure S12). Finally, given the coupling between transcription and translation in bacteria (Burmann et al., 2010; Proshkin et al., 2010), pausing at SD sites could be exploited for transcriptional regulation. We observed internal SD sites and pausing near the stop codon of transcription attenuation leader peptides (Kolter and Yanofsky, 1982), including *trpL* and *thrL* (Figure S13). In contrast to ribosome stalling at regulatory codons during starvation, slow translation near the stop codon could protect alternative structural mRNA elements to prevent the formation of anti-termination stem loops, thereby ensuring proper transcription termination (Elf and Ehrenberg, 2005). Our approach and the genome-wide data lay the groundwork for further gene specific functional studies of translational pausing.

Perspective

From a more practical perspective, ribosome pausing at internal SD sites presents both a challenge and an opportunity for heterologous protein expression in bacteria. Overexpression of eukaryotic proteins with strong internal SD sites would sequester ribosomes and compromise protein yield. Internal SD sequences could be reduced by recoding the gene, which has not been considered in conventional strategies

of simple codon optimization or overexpression of rare tRNAs. Conversely, recoding can introduce internal SD sites if pausing is required for cotranslational processing. Positioning of internal SD sites therefore adds another dimension to the optimization of heterologous protein expression.

Experimental Procedures

E. coli MG1655 and *B. subtilis* 168 were used as wild-type strains. *E. coli* BJW9 has synonymous substitutions at G141 and G142 in the *ompF* gene. The orthogonal ribosome experiment was performed in *E. coli* BW25113 with two plasmids: pSC101-G9, expressing orthogonal 16S rRNA, and pJW1422, expressing O-SD-*lacZ* mRNA. pSC101-G9 was a gift from J. Chin (Rackham and Chin, 2005). pJW1422 has *lacZ* driven from a *tacII* promoter and an O-ribosome binding site 5'-AUCCCA-3'. Luria broth was used for *B. subtilis* culture. Cell cultures were harvested at a D_{600} of 0.3–0.4. Flash freezing and ribosome footprinting was previously described (Oh et al., 2011). 3 mM of 5'-guanylyl imidodiphosphate was added to the lysate before thawing and during footprinting to prevent translation after lysis. Conversion of mRNA footprints to a complementary DNA library was previously described (Ingolia et al., 2011; Oh et al., 2011). Deep sequencing was performed on an Illumina HiSeq 2000 system, and the results were aligned to reference genomes using Bowtie v. 0.12.0. The cross-correlation function is defined as:

$$C_i = \frac{\left\langle x_j +_i y_j \right\rangle - \mu_X \mu_Y}{\sigma_X \sigma_Y}$$

for the series ($X = x_1, x_2, ..., x_N$) and ($Y = y_1, y_2, ..., y_N$). where μ_X and $_X$ are respectively the average and standard deviation of series X.

Acknowledgements

We thank E. Reuman, D. Burkhardt, C. Jan, C. Gross, J. Elf, and members of the Weissman laboratory for discussions; J. Dunn for ribosome profiling data on S. cerevisiae; C. Chu for help with sequencing; and J. Chin for orthogonal ribosome reagents and advice. This research was supported by the Helen Hay Whitney

Foundation (to G.W.L.) and by the Howard Hughes Medical Institute (to J.S.W.).

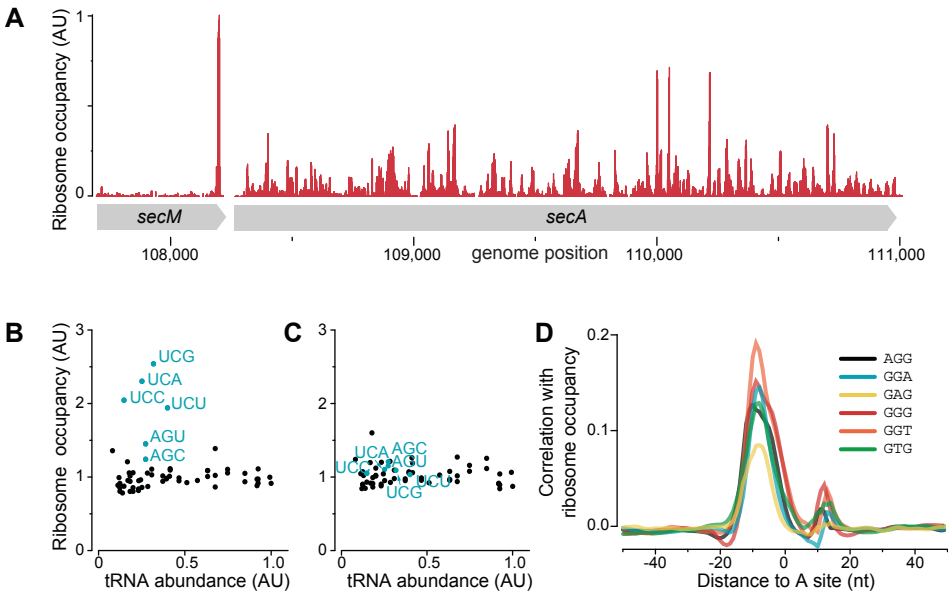
Author Contributions

G.W.L. and J.S.W. designed the experiments. G.W.L. performed experiments and analysed the data. E.O. provided technical support and preliminary data. G.W.L. and J.S.W. wrote the manuscript.

Author Information

The footprint sequencing data are deposited in the Gene Expression Omnibus (GEO) under accession number GSE35641. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S.W. (weissman@cmp.ucsf.edu).

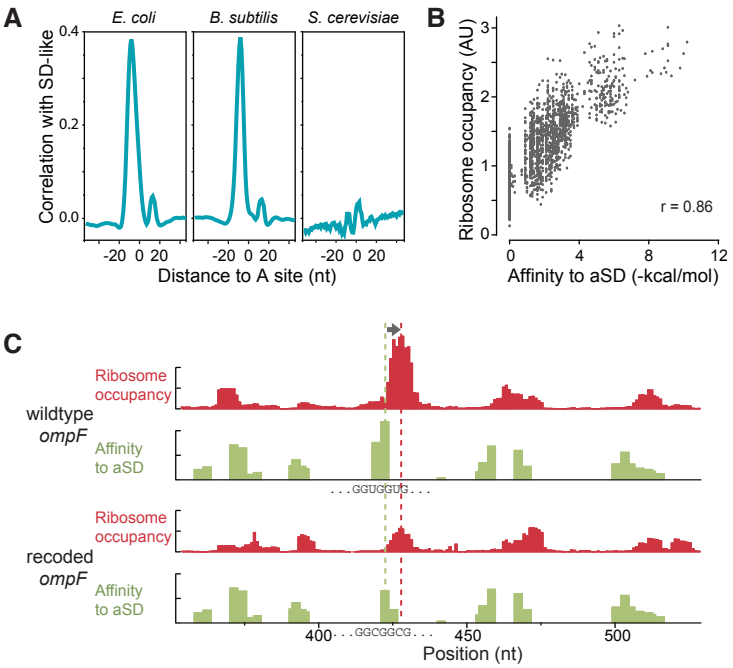
Figure 1



Analysis of Translational Pausing Using Ribosome Profiling in Bacteria

(A) Validation of the ribosome stalling site in the *secM* mRNA. **(B) (C)** Average ribosome occupancy of each codon relative to their respective tRNA abundances for *E. coli*. **(B)** For growth in Luria broth, elevated occupancy at serine codons (blue) probably reflects preferential depletion of this amino acid. **(C)** In glucose-rich medium, the ribosome occupancy is independent of tRNA abundance. **(D)** Plot of cross-correlation function between ribosome occupancy profiles and the presence of the indicated trinucleotide sequences for *E. coli*.

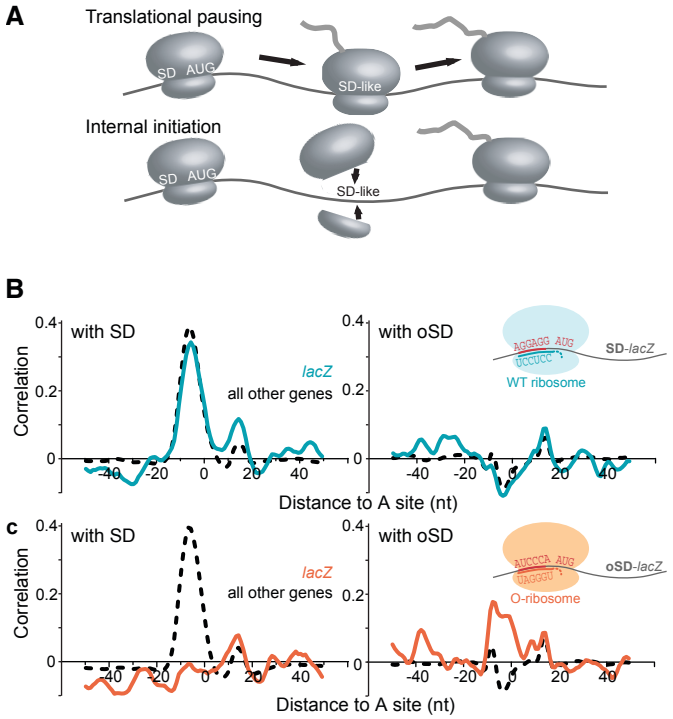
Figure 2



Relationship between Ribosome Pausing and Internal Shine-Dalgarno Sequences

(A) Correlation between ribosome occupancy and SD-like features for *E. coli*, *B. subtilis*, and *S. cerevisiae*. **(B)** Plot of the average ribosome occupancy of hexanucleotide sequences relative to their affinity for the anti-Shine-Dalgarno sequence. **(C)** Reprogrammed pausing by recoding the *ompF* mRNA. Ribosome occupancy (red) increases when the A-site is 8–11 nucleotides downstream (arrow) of SD-like features (green). Synonymous mutations replacing the SD-like sequence (GGUGGUG) in wild-type *ompF* (top) with a sequence (GGCGGCG) with lower affinity for the aSD site (bottom) caused a corresponding decrease in ribosome pausing.

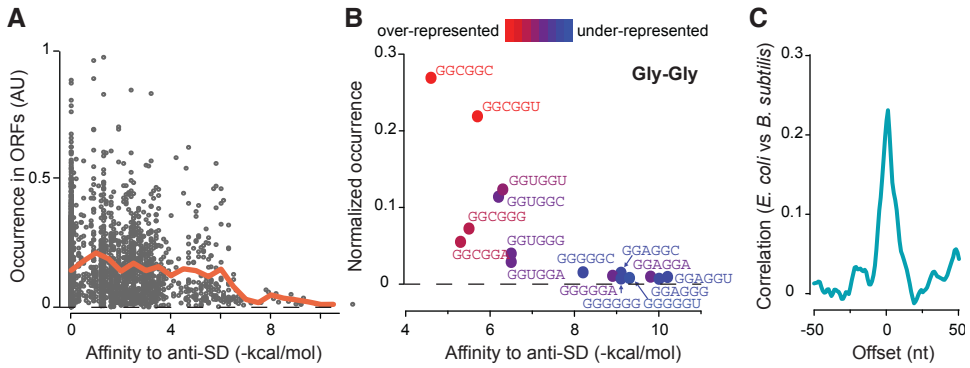
Figure 3



Pausing of Elongating Ribosomes due to SD–aSD Interaction

(A) Diagram of two models that could account for the excess ribosome density at internal SD-like sequences. **(B)** Ribosome occupancy of *lacZ* mRNA translated by wild-type ribosome. Like other genes translated by the wild-type ribosome, the ribosome occupancy pattern along *lacZ* is correlated with the presence of SD-like sequences (left), not with the O-SD sequence (right). Cyan, *lacZ*; black, all other genes. **(D)** Ribosome occupancy of *lacZ* mRNA translated by orthogonal ribosomes (O-ribosome). Unlike other genes in the same cells, the specialized O-SD *lacZ* has ribosome pausing at internal O-SD-like sequences (right), not at SD-like sites (left). Orange, *lacZ*; black, all other genes.

Figure 4



Selection against SD-like Sequences and the Constraint on Protein Coding

(A) Rate of occurrence of hexanucleotide sequences in *E. coli* messages relative to their predicted affinity for the aSD site. The orange line shows the average occurrence within a bin size of 0.5 kcal mol^{-1}. **(B)** Occurrence of codon pairs for Gly-Gly residues relative to their predicted affinity for the aSD site. The color coding represents the enrichment in occurrence of codon pairs after correcting for the usage of single codons. **(C)** Cross-correlation function of ribosome occupancy profiles between conserved genes in *E. coli* and *B. subtilis*. Zero offset means that the two sequences are aligned at each amino acid residue.

EXTENDED EXPERIMENTAL PROCEDURES

Strains, Plasmids, and Growth Conditions

E. coli K-12 MG1655 and *B. subtilis* 168 were used as wild-type strains. *E coli.* BJW9 contains recoded *ompF* (GGT to GGC synonymous substitutions at G141 and G142) that was introduced in MG1655 by lambda-Red recombination (Datsenko and Wanner, 2000) at the endogenous locus. Orthogonal ribosome experiments were performed in *E. coli* BW25113, which is a K-12-derived strain with a *lacZ* deletion (Datsenko and Wanner, 2000).

pSC101-G9 expresses orthogonal 16S rRNA from an intact *rrnB* operon except that the 3' end of *rrsB*, which codes for the 16S rRNA, is changed from 5'-CCTCCTTA-3' to 5'-TGGGATTA-3' (Rackham and Chin, 2005). pJW1422 harbors the *lacZ* gene under the *tacII* promoter. The ribosome-binding site of the *lacZ* mRNA is replaced with 5'-AUCCCA-3', allowing for the initiation of translation by orthogonal ribosomes.

Unless otherwise noted, *E. coli* strains were grown in MOPS media supplemented with 0.2% glucose, 20 amino acids, vitamins, bases, and micronutrients (Neidhardt et al., 1974) (Teknova). *B. subtilis* was grown in Luria broth (BD Difco). For strains containing pSC101-G9 and pJW1442, media was supplemented with 25 μg ml^{-1} kanamycin and 15 μg ml^{-1} tetracycline. For experiments with *E. coli*, an overnight liquid culture was diluted 1:400 into fresh medium. For experiments with *B. subtilis*, an overnight culture on an LB plate was washed and diluted to a D_{600} of 0.00125 in Luria broth. Cell cultures were grown at 37°C until the D_{600} reached 0.3–0.4.

Ribosome Profiling

The protocol for bacterial ribosome profiling with flash freezing was previously

described (Oh et al., 2011). 200 ml of cell culture was rapidly filtered through a pre-warmed nitrocellulose filter with a 200-nm pore size. Cell pellet was flash frozen in liquid nitrogen and combined with 650 μl of frozen lysis buffer (10 mM $MgCl_2$, 100 mM NH_4Cl, 20 mM Tris-HCl pH 8.0, 0.1% Nonidet P40, 0.4% Triton X-100, 100 U ml^{-1} DNase I (Roche), 3 mM 5'-guanylyl imidodiphosphate (GMPPNP), 1 mM chloramphenicol). Addition of GMPPNP, together with chloramphenicol, inhibits translation after lysis. Cells were pulverized in 10-ml canisters pre-chilled in liquid nitrogen. Lysate containing 0.5 mg of RNA was digested for 1 hour with 750 U of micrococcal nuclease (Roche) at 25°C. The ribosome-protected fragments were isolated using a sucrose gradient followed by phenol extraction. The footprints were ligated to a 5'-adenylated and 3'- blocked DNA oligonucleotide (/5rApp/CTGTAGGCACCATCAAT/3ddc, Integrated DNA Technologies). Unless otherwise noted, the ligation was performed with truncated T4 RNA ligase 2 (New England Biolabs) as previously described (Ingolia et al., 2011; Oh et al., 2011). To remove lot-to-lot variability in the activity from the commercial source, we have recently switched to recombinantly expressed truncated T4 RNA ligase 2 K227Q produced in our laboratory. We used this ligase to generate a library for the high-coverage data set for *E. coli*. The 3'-ligated RNA fragments were converted to a sequenceable DNA library by using reverse transcription, circularization, and PCR amplification as previously described (Ingolia et al., 2011; Oh et al., 2011).

Sequencing was performed on an Illumina HiSeq 2000 system. Sequence alignment with Bowtie v.0.12.0 mapped the footprint data to the reference genomes NC_000913.fna (*E. coli*) or NC_000964.fna (*B. subtilis*) obtained from the NCBI Reference Sequence Bank. Data from BJW9 was aligned to a reference modified from NC_000913.fna. The footprint reads varied between 25 and 42 nucleotides in length, mostly as a result of the specificity of micrococcal nuclease. In contrast to eukaryotic

systems, in which the 5' end of the footprint is sufficient to carry the positional information (Ingolia et al., 2009; Ingolia et al., 2011), we distributed the positional information into several nucleotides across the center of the footprint (Oh et al., 2011). For each footprint read, the center residues that were at least 12 nucleotides from either end were given the same score, which was weighted by the length of the fragment.

To assign the A-site position to the center of ribosome footprints, we made use of the ribosome density at two independent sets of well-defined pause sites. The first set consisted of pausing at stop codons (Oh et al., 2011), where the ribosomal A-site was aligned to stop codons before binding of release factors. The second set consisted of peptide-mediated ribosome stalling sites, where the A-site codons had been identified. These two alignments were consistent with each other. In addition, the pausing at serine codons at the A-site during starvation confirmed the position assignment of ribosome footprints.

mRNA Sequencing

Total RNA was phenol extracted from the same lysate that was used for ribosome footprinting (Oh et al., 2011). Ribosomal RNA and small RNA were respectively removed from the total RNA with MICROBExpress and MEGAclear (Ambion). mRNA was randomly fragmented as previously described (Ingolia et al., 2009). The fragmented mRNA sample was converted to a complementary DNA library with the same strategy as done for ribosome footprints, and was described previously (Oh et al., 2011).

Data Analysis

Data analysis was performed with scripts written for Python 2.6.6. Global pausing

analyses were based on 2,257 genes from *E. coli* and 1,580 genes from *B. subtilis*, with an average coverage of at least ten sequencing reads per codon in the ribosome profiling data set. In addition, analyses on 997 genes from *E. coli* and 1,189 genes from *B. subtilis*, with an average coverage of between one and ten sequencing reads per codon, showed qualitatively consistent results. For *E. coli*, *tufA* and *tufB* genes were not included in the analysis because of their sequence homology with each other. Genes with known frame-shifting sites (*prfB* and *dnaX*) were not included in codon-specific analyses. On gene-specific analyses, the coverage was at least 30 sequencing reads per codon in each case.

To focus on the kinetics of translation elongation, the analysis was performed on the basis of ribosome occupancy profiles within protein-coding genes, excluding the first and last ten codons. To calculate the average ribosome occupancy associated with each codon at the A-site, the ribosome occupancy profile of each gene was normalized by the mean occupancy of the gene, and the normalized occupancy for each codon was averaged across all genes. Similarly, the average ribosome occupancy for each hexanucleotide at the SD position was calculated by averaging normalized occupancy between 7 and 12 nucleotides downstream of the hexanucleotide sequence. For each codon, the corresponding tRNA abundance, plotted in Figures 1 and S7, was the sum of the expression levels of the cognate tRNA species previously measured (Dong et al., 1996; Kanaya et al., 1999).

To identify dominating sequence features either upstream or downstream of the pausing sites, we slid the ribosome occupancy profile ($X = x_1, x_2, ..., x_N$) along the coding sequence, and, at every offset position i, calculated the correlation with the presence of a given sequence ($Y = y_1, y_2, ..., y_N$). In mathematical terms, it is given by the normalized

cross-correlation function (C_i):

$$C_i = \frac{\left\langle x_j +_i y_j \right\rangle - \mu_X \mu_Y}{\sigma_X \sigma_Y}$$

where μ_X and μ_Y are the average of the series X and Y, respectively. $_X$ and $_Y$ are the standard deviations of the series X and Y, respectively. $\langle x_j +_i y_j \rangle$ is the expectation value of $x_j +_i y_j$ for all possible values of j. We used Python to calculate $\sum x_j +_i y_j$ using the *correlate* function in the *same* mode of the numpy package. The expectation value is obtained by dividing the summation by $N - |i|$. For each gene with more than ten sequencing reads per codon and longer than 160 base pairs, we calculated the normalized cross-correlation function. The average over these cross-correlation functions is presented in this paper.

Hybridization Free-Energy Prediction

The hybridization free energy between mRNA and the aSD site was predicted with the RNAsubopt program in the Vienna RNA package (Gruber et al., 2008). The energy was predicted for 37°C with a contribution from dangling ends. For each hexanucleotide sequence, the lowest possible hybridization free energy was assigned as its affinity for the aSD site. We used the eight-nucleotide sequence 5'-CACCUCCU-3' as the aSD sequence. To calculate the cross-correlation function between ribosome occupancy profile and SD-like features (Figure 2A), we built the aSD affinity profile for each mRNA by scanning the transcript in overlapping units of ten nucleotides and calculating the affinity of aSD to the middle eight nucleotides. The affinity was assigned to the eighth position in the ten-nucleotide window, which corresponds to U in the canonical SD sequence. The distance from the P-site to U in the canonical SD sequence is often defined as the aligned spacing (Chen et al., 1994). Because we align ribosome

footprints to the A-site, the distance reported is three nucleotides longer than the aligned spacing.

Analysis of O-ribosome Translated Messages

Because a *lacZ* message whose translation is driven by O-SD is exclusively translated by O-ribosomes (Rackham and Chin, 2005), the translational pausing model outlined in Figure 3A predicts that for the O-SD driven *lacZ*, there will be both the appearance of new ribosome density peaks at internal sites that resemble the O-SD sequence and the disappearance of peaks at the SD-like sequences found when translation is driven by the wild-type SD sequence. This prediction is confirmed by our data in Figure 3C in which the ribosome occupancy profile of *lacZ* with O-SD-driven translation no longer shows a correlation with SD-like sequences, and is instead correlated with O-SD-like sequences. Moreover, because the endogenous messages are still translated solely by wild-type ribosomes even when the O-ribosome is present, the ribosome peaks in the endogenous messages are found at SD-like sequences, not at sequences that resemble the O-SD site, regardless of whether O-ribosomes are present. This is again confirmed by the data shown in Figure 3C.

Conservation Analysis

Conservation analysis of pausing patterns in *E. coli* and *B. subtilis* was performed in a set of 31 proteins from the curated alignment database AMPHORA (Wu and Eisen, 2008). The nucleotide sequences and the ribosome density profiles were trimmed and concatenated on the basis of the protein alignment. The cross-correlation function between *E. coli* and *B. subtilis* ribosome occupancy was calculated for each gene, and then averaged over 31 genes to give the conservation of pausing patterns.

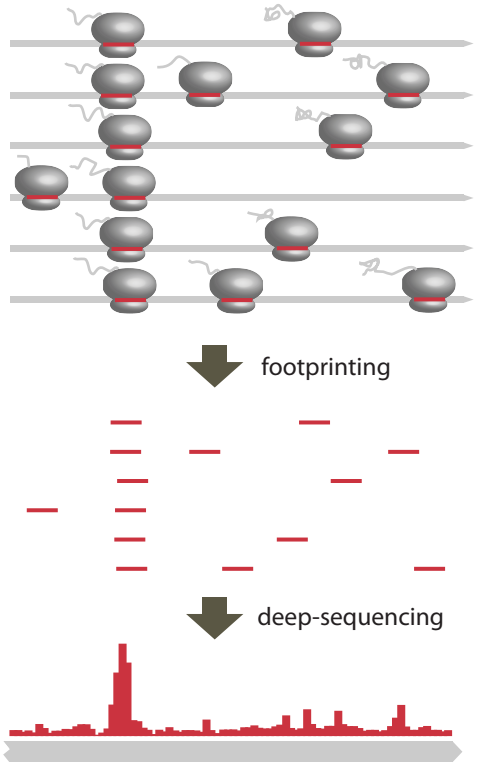
Occurrence of Hexamers and Codon Pairs

The occurrence of hexamers and codon pairs was counted from annotated protein-coding genes available from the NCBI Reference Sequence Bank. Normalized occurrence ($p_{i,j}$) was calculated by dividing the occurrence of a given codon pair (i and j) by the total occurrence of the corresponding amino acid pair. The correction for the usage of single codons was calculated by dividing the normalized occurrence of the codon pair ($p_{i,j}$) by the frequency of the two individual codons (q_i and q_j), whereby enrichment = $p_{i,j} / q_i q_j$. The frequency of individual codons was normalized to the occurrence of the corresponding amino acid.

Protein Structure Analysis

Protein secondary structure was predicted by the PSIPRED method (Jones, 1999), with the filtered reference database UniRef90 (Li et al., 2001). Secondary structures were predicted for 271 proteins. Cross-correlation function between the structural assignment with either ribosome occupancy or SD-like features was calculated at the nucleotide level.

Figure S1



Schematic of Ribosome Profiling

The protocol for bacterial ribosome profiling with flash freezing was previously described (Oh et al., 2011). Polysome-containing cell lysate was treated with micrococcal nuclease (MNase) to generate ribosome-protected mRNA fragments. The mRNA footprints were converted into a sequenceable DNA library. Regions of mRNA that have higher ribosome occupancy give rise to more sequenced fragments.

Figure S2

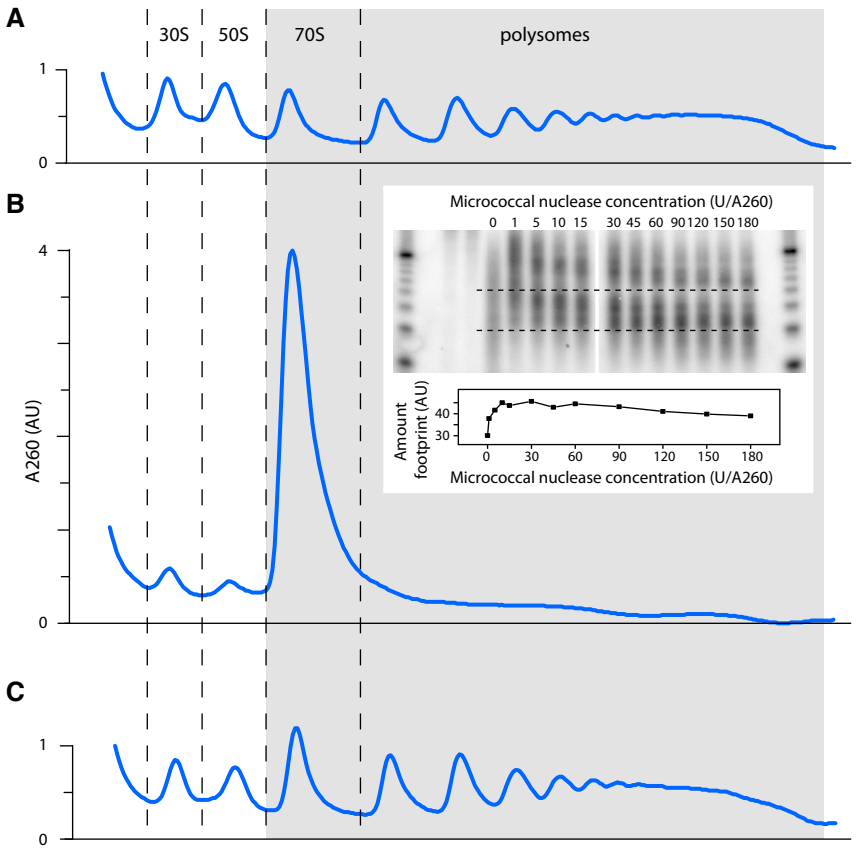
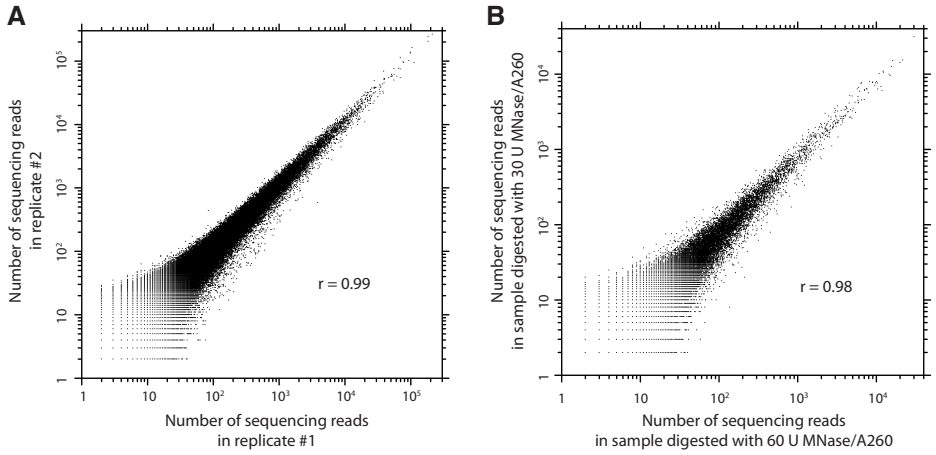

A
30S
50S
70S
polysomes
1
0
B
4
0
A260 (AU)
Micrococcal nuclease concentration (U/A260)
0 1 5 10 15 30 45 60 90 120 150 180
Amount footprint (AU)
40
30
0 30 60 90 120 150 180
Micrococcal nuclease concentration (U/A260)
C
1
0

Figure S2

Polysome Profiles of Ribosome Footprinting for *E. coli*

(A) Polysome profile of flash frozen and pulverized cell lysate. ~77% of total RNA was in assembled ribosomes (shaded area), of which ~87% was in the polysome fraction. **(B)** Polysome profile after treatment with micrococcal nuclease at 25°C for one hour. The amount of RNA in the assembled ribosome (shaded area) was the same as that of the undigested lysate, indicating that assembled ribosomes stayed intact during footprinting. Consistent with this observation, nuclease protection assay (inset) showed a constant level of ribosome-protected footprints over a range of MNase concentrations. Footprint counts (between dashed lines) were plotted as a function of nuclease concentration. In the ribosome profiling experiments, we used 60 U of nuclease per 1 A260 unit of RNA. Nuclease protection assay was previously described (Ingolia et al., 2009) using the mirVana miRNA detection kit. $^{[\alpha 32P]}$UTP labeled antisense probe for *gapA* from *E. coli* was generated using the MAXIscript kit with T7 RNA polymerase. **(C)** Polysome profile after incubation at 25°C for one hour without MNase. The fraction of RNA in the assembled ribosome (shaded area) was again the same as that of the undigested lysate. The ratio of 70S particles to polysome fractions increased during the 25°C incubation, which is likely due to breakage of mRNA in between ribosomes.

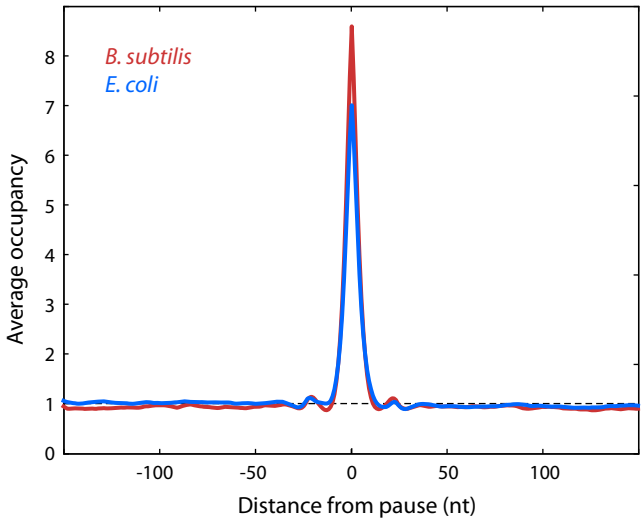
Figure S3



Reproducibility of Bacterial Ribosome Profiling

(A) Reproducibility among biological replicates. Each dot corresponds to the number of sequencing reads mapped to a particular position on mRNA in ribosome profiling experiments from two separate cultures of *E. coli*. The Pearson correlation coefficient is 0.99. **(B)** Effect of MNase concentration. Lysate of *B. subtilis* was treated with either 60 U or 30 U of MNase per 1 A260 unit of RNA, and the number of ribosome protected fragments at each position on mRNA were compared. The Pearson correlation coefficient of 0.98, confirming that nuclease digestion introduces negligible bias at the working concentration of MNase. In addition, the correlation between Shine-Dalgarno-like sequences and pausing was unaffected by the 2-fold change in the amount of MNase.

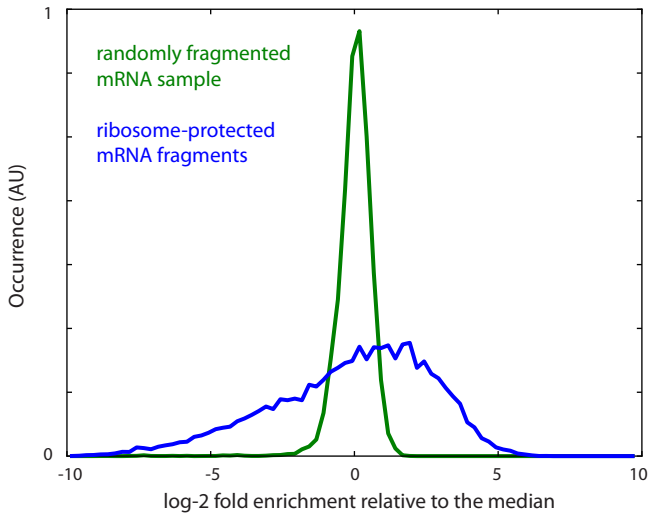
Figure S4



Average Ribosome Density before and after Translational Pausing Sites

Pauses with ribosome occupancy 5-fold greater than the mean were aligned at position 0. The ribosome occupancy surrounding a pause was normalized by the mean occupancy of the message, and averaged over all pausing sites. We observe no loss of ribosome density immediately before and after pauses, indicating that translation within coding sequences is a continuous process with negligible internal initiation and early termination at the pausing sites. Furthermore, this observation also argues that there is negligible ribosome movement after cells were flash frozen, which would lead to depletion of ribosome density after pausing sites.

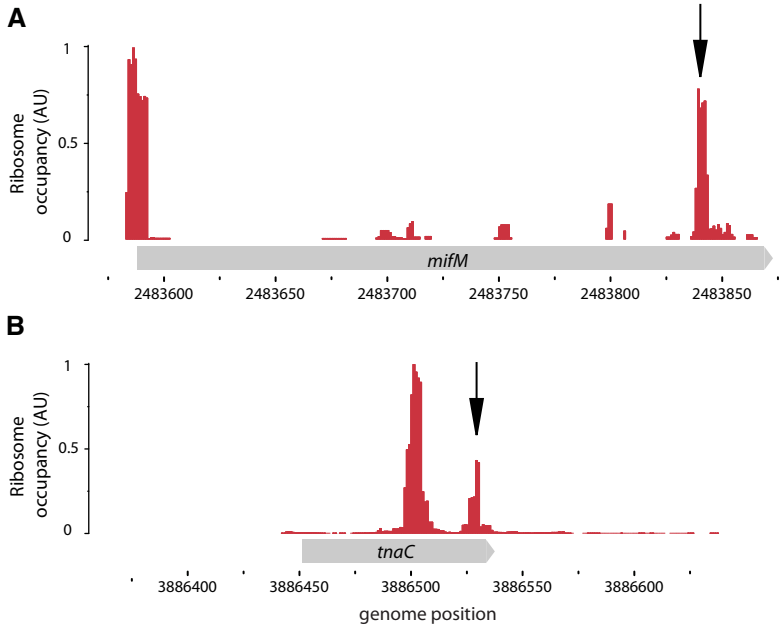
Figure S5



Variation of Ribosome Occupancy and Randomly Fragmented mRNA Sample for *E. coli*

Ribosome footprints (blue) and randomly fragmented mRNA (green) were converted to sequenceable DNA libraries using the same protocol. The frequency of sequencing reads at each codon on each message was normalized to the median frequency of the message. Histograms of log[2] enrichment relative to the median were plotted for codons in the genes that have at least 10 sequencing reads per codon on average. Ribosome occupancy exhibited greater variations than that introduced during the conversion of RNA fragments into a sequenceable DNA library.

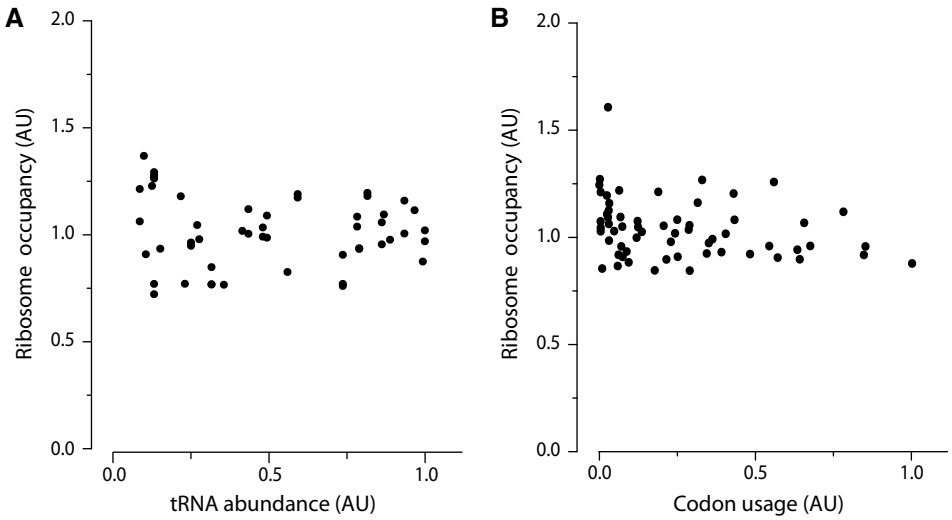
Figure S6



Ribosome Occupancy Profile of Genes with Translational Stalling Sites

(A) The *mifM* gene in *B. subtilis*. **(B)** The *tnaC* gene in *E. coli*. The arrows point at the position of known stalling sites. In *tnaC* we observed a second ribosome queuing ~30 nt upstream the known stalling site. The presence of this second ribosome immediately before the stalling site would be difficult to detect using conventional assays based on primer extension. It is plausible that a trailing ribosome is queuing behind the ribosome that is stalled downstream.

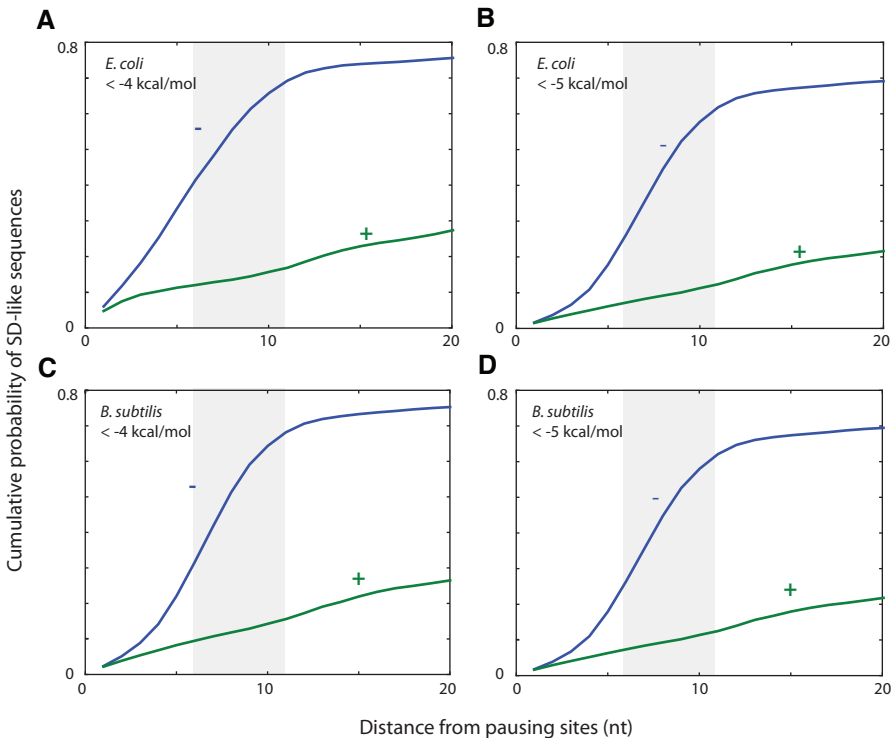
Figure S7



Average Ribosome Occupancy of Codons

(A) Ribosome occupancy relative to the corresponding tRNA abundance in *B. subtilis*. Similar to Figures 1C and 1D, average ribosome occupancy of each codon is plotted relative to their respective tRNA abundance. Codons with undetermined tRNA abundance were not included. The codon-specific ribosome occupancy was uncorrelated with the tRNA abundance. **(B)** Ribosome occupancy relative to codon usage in *E. coli*. The codon usage was calculated from a group of 321 highly expression genes that have at least 500 sequencing reads per codon on average in the dataset. The average ribosome occupancy was calculated from 2,255 genes with at least 10 sequencing reads per codon.

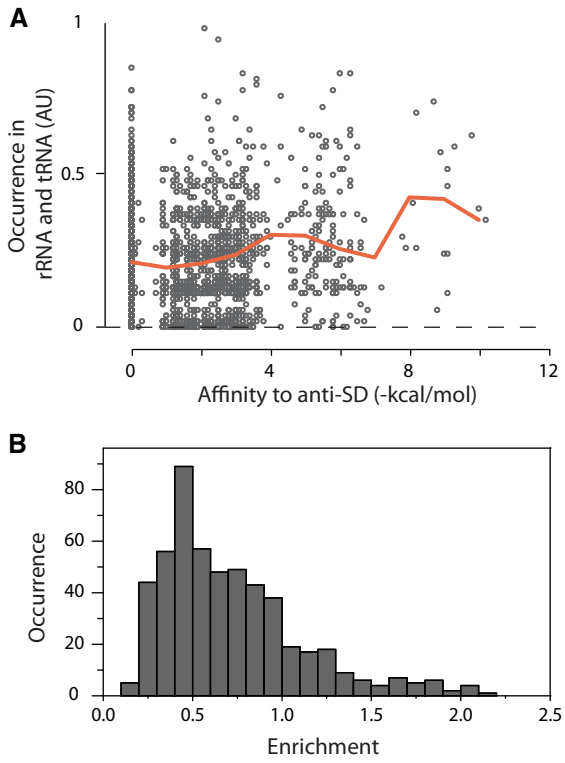
Figure S8



Fraction of Pauses Associated with SD-like Sequences

Cumulative probability of having SD-like sequences either upstream (–) or downstream (+) from pausing sites was plotted against the distance from the pausing sites in **(A, B)** *E. coli* and **(C, D)** *B. subtilis*. The cumulative probability is the probability of having at least one SD-like sequence within a certain distance from the pausing site. SD-like sequences were defined as hexamer sequences with affinity to aSD **(A, C)** less than –4 kcal/mol or **(B, D)** less than –5 kcal/mol. Pausing sites with ribosome occupancy greater than 10-fold of the mean (~2 pauses/gene) were included in this analysis. ~70% of the pauses were associated with SD-like sequences upstream (shaded).

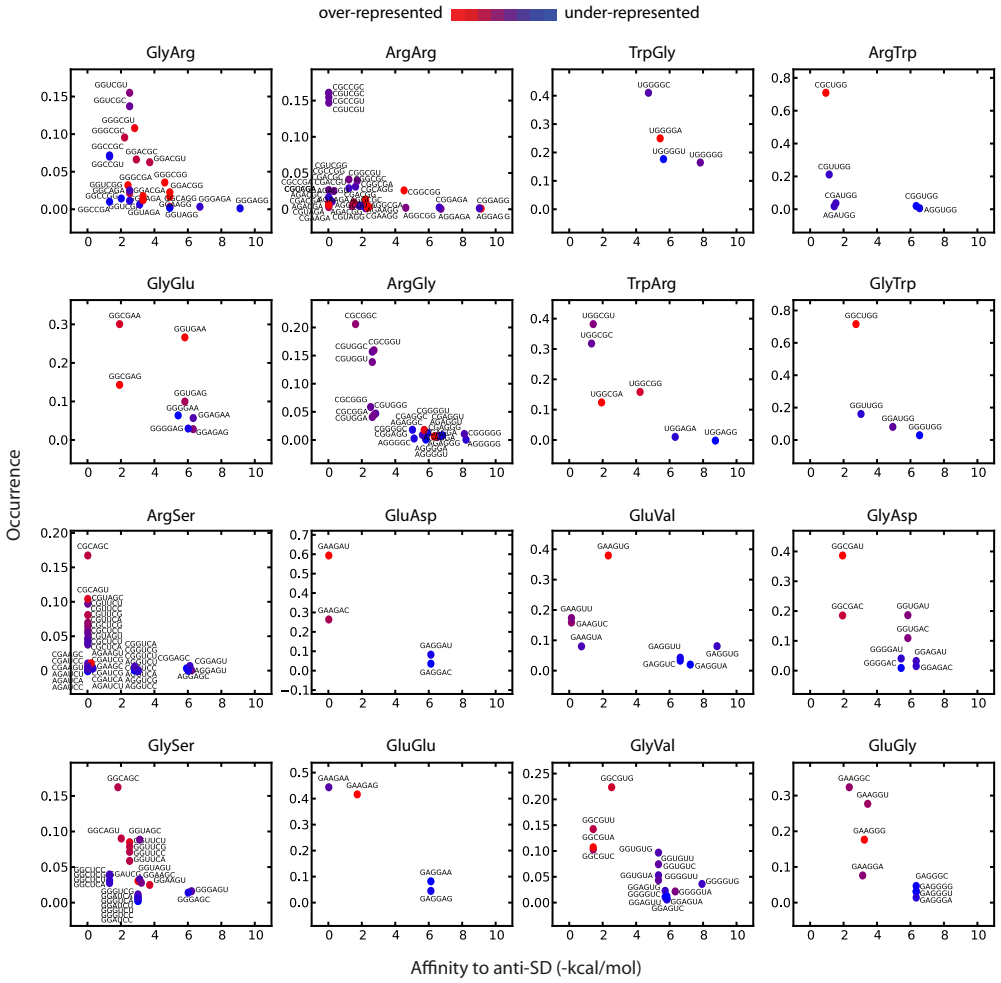
Figure S9



Occurrence of SD-like Sequences

(A) The occurrence of hexamer sequences in rRNA and tRNA relative to the affinity to anti-SD in *E. coli*. The orange line shows the average occurrence within a bin size of 0.5 kcal/mol. Unlike hexamers in protein coding sequences, strong SD-like hexamers were not avoided. **(B)** Histogram of enrichment of internal SD-like sequences in the mRNA of 533 bacterial species in the AMPHORA35 database. The enrichment level of each species was calculated based on its GC content. SD-like hexamers (with predicted hybridization energy less than –7 kcal/mol) were avoided in the majority of bacterial species. The avoidance of SD-like sequences is one of many forces, including GC bias and mutational bias, that determine the genome composition.

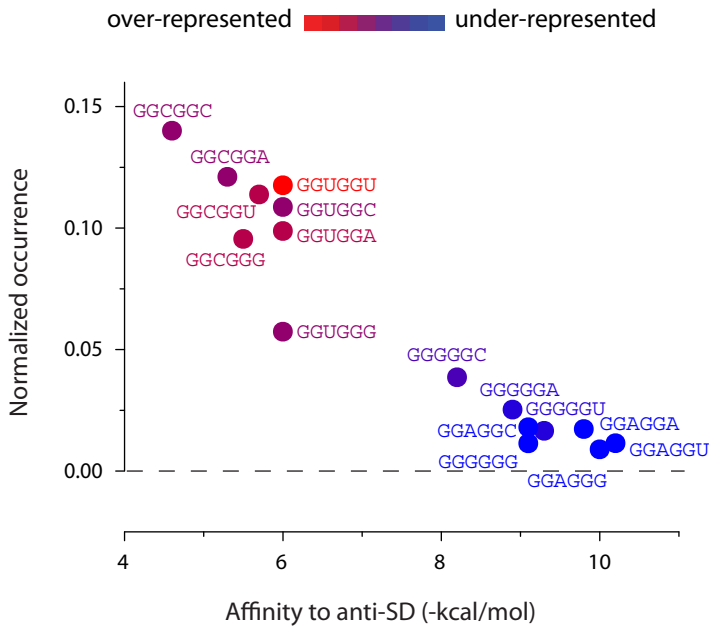
Figure S10



Disenrichment of Codon Pairs that Resemble SD Sequences in *E. coli*

We calculated the normalized occurrence of codon pairs (y-axes) and the enrichment relative to single codon usage (color coded) for 16 pairs of amino acids that can be coded with SD sites (less than –6 kcal/mol). The occurrence was normalized within each group of codon pairs encoding the same pairs of amino acids. Similar to Gly-Gly pairs, strong SD-like codon pairs appear less often than what is expected from the single codon usage.

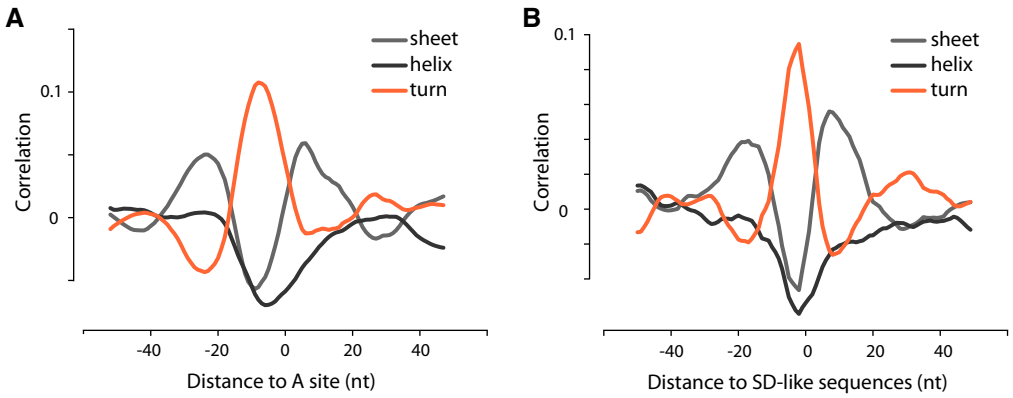
Figure S11



Occurrence of GGNGGN Sequences that Are Not Aligned to Gly-Gly pairs in *E. coli* Protein Coding Sequences

The occurrence of GGNGGN that does not encode two glycine codons was plotted against the affinity to the anti-SD site. The color coding represents the enrichment in occurrence after correcting for the usage of single trinucleotide sequence. The fact that the same trend exists regardless of reading frame information supports the notion that the preference of codon pairs stems from properties of the sequence, rather than properties of the codon or the tRNA.

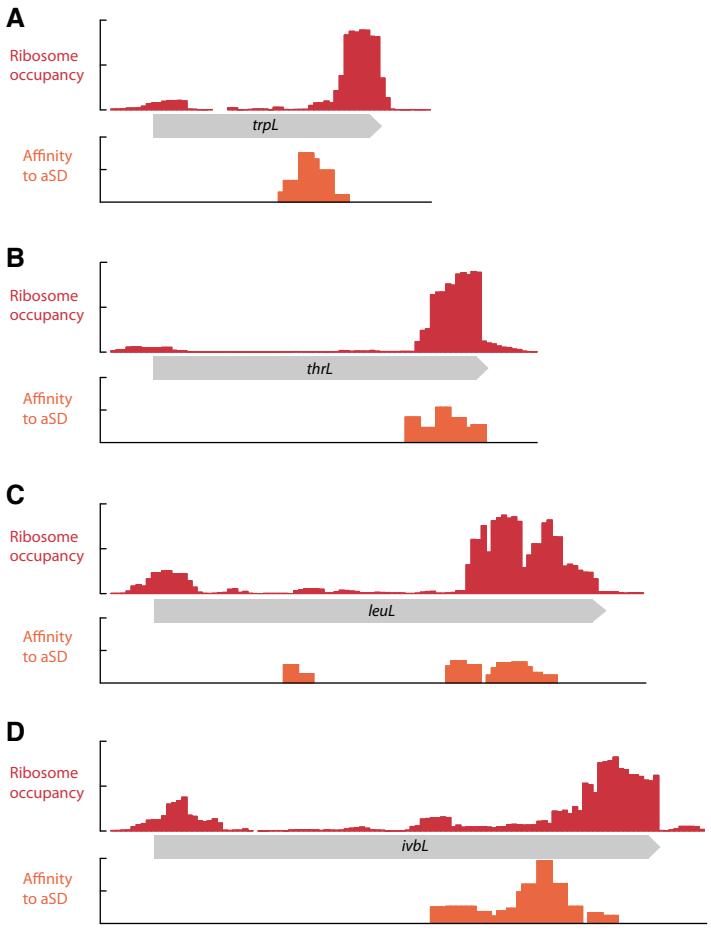
Figure S12



Correspondence of Protein Structure and Ribosome Pausing

(A) Correlation between protein secondary structures and ribosome occupancy profiles. Translational pauses were over-represented in places where the newly synthesized polypeptides correspond to turns in a protein. **(B)** Correlation between protein secondary structures and SD-like sequences. SD-like sequences are also over-represented in regions encoding protein turns. The pause sites, including at most Gly-Gly residues, are over-represented in protein turns and unstructured regions. Therefore pausing could potentially facilitate independent folding of adjacent structural motifs. An important caveat is that pausing at SD sites may occurs when the amino acid residues translated from SD sites, such as Gly-Gly, are still within the ribosome exit tunnel. Whether the structured region would be outside the exit tunnel therefore needs to be determined on a case-by-case basis.

Figure S13



Ribosome Pausing near the End of Leader Peptide Sequences of Amino Acid Biosynthesis Operons

Ribosome density is low at the beginning of leader sequences and high near the end. Slow translation near the stop codon may provide additional protection for the structural mRNA elements to promote transcription termination.

REFERENCES

Andersson, S.G., and Kurland, C.G. (1990). Codon preferences in free-living microorganisms. Microbiol. Rev. *54*, 198–210.

Baranov, P.V., Gesteland, R.F., and Atkins, J.F. (2002). Release factor 2 frameshifting sites in different bacteria. EMBO Rep. *3*, 373–377.

Burmann, B.M., Schweimer, K., Luo, X., Wahl, M.C., Stitt, B.L., Gottesman, M.E., and Rösch, P. (2010). A NusE:NusG complex links transcription and translation. Science *328*, 501–504.

Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. Nucleic Acids Res. *22*, 4953–4957.

Chiba, S., Kanamori, T., Ueda, T., Akiyama, Y., Pogliano, K., and Ito, K. (2011). Recruitment of a species-specific translational arrest module to monitor different cellular processes. Proc. Natl. Acad. Sci. USA *108*, 6073–6078.

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

Dong, H., Nilsson, L., and Kurland, C.G. (1996). Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. J. Mol. Biol. *260*, 649–663.

Elf, J., and Ehrenberg, M. (2005). What makes ribosome-mediated transcriptional attenuation sensitive to amino acid limitation? PLoS Comput. Biol. *1*, e2.

Gong, F., and Yanofsky, C. (2002). Instruction of translating ribosome by nascent peptide. Science *297*, 1864–1867.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36*, W70–74.

Gutman, G.A., and Hatfield, G.W. (1989). Nonrandom utilization of codon pairs in Escherichia coli. Proc. Natl. Acad. Sci. USA *86*, 3699–3703.

Hui, A., and de Boer, H.A. (1987). Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in Escherichia coli. Proc. Natl. Acad. Sci. USA *84*, 4762–4766.

Ikemura, T. (1981). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J. Mol. Biol. *151*, 389–409.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell *147*, 789–802.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. *292*, 195–202.

Kanaya, S., Yamada, Y., Kudo, Y., and Ikemura, T. (1999). Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene *238*, 143–155.

Kolter, R., and Yanofsky, C. (1982). Attenuation in amino acid biosynthetic operons. Annu. Rev. Genet. *16*, 113–134.

Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat. Struct. Mol. Biol. *16*, 589–597.

Larsen, B., Wills, N.M., Gesteland, R.F., and Atkins, J.F. (1994). rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. J. Bacteriol. *176*, 6842–6851.

Li, W., Jaroszewski, L., and Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics *17*, 282–283.

Nakatogawa, H., and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. Cell *108*, 629–636.

Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. J. Bacteriol. *119*, 736–747.

Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G.*, et al.* (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. Cell *147*, 1295–1308.

Pedersen, S. (1984). Escherichia coli ribosomes translate in vivo with variable rate. EMBO J. *3*, 2895–2898.

Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. *12*, 32–42.

Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science *328*, 504–508.

Pruss, B.M., Nelms, J.M., Park, C., and Wolfe, A.J. (1994). Mutations in NADH:ubiquinone oxidoreductase of Escherichia coli affect growth on mixed amino acids. J. Bacteriol. *176*, 2143–2150.

Rackham, O., and Chin, J.W. (2005). A network of orthogonal ribosome x mRNA pairs. Nat. Chem. Biol. *1*, 159–166.

Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007). Escherichia coli physiology in Luria-
Bertani broth. J. Bacteriol. *189*, 8746–8749.

Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA *71*, 1342–1346.

Sorensen, M.A., Kurland, C.G., and Pedersen, S. (1989). Codon usage determines translation rate in Escherichia coli. J. Mol. Biol. *207*, 365–377.

Sorensen, M.A., and Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in Escherichia coli. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. J. Mol. Biol. *222*, 265–280.

Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol. *180*, 549–576.

Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F., and Gesteland, R.F. (1988). Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in Escherichia coli. EMBO J. *7*, 1503–1507.

Wen, J.D., Lancaster, L., Hodges, C., Zeri, A.C., Yoshimura, S.H., Noller, H.F., Bustamante, C., and Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. Nature *452*, 598–603.

Wu, M., and Eisen, J.A. (2008). A simple, fast, and accurate method of phylogenomic inference. Genome Biol. *9*, R151.

CHAPTER FOUR

Investigating the Cellular Function of Trigger Factor by Selective Ribosome Profiling

Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo

Eugene Oh,[1,2,3,6] Annemarie H. Becker,[5,6] Arzu Sandikci,[5] Damon Huber,[5] Rachna Chaba,[4] Felix Gloge,[5] Robert J. Nichols,[4] Athanasios Typas,[4] Carol A. Gross,[4] Günter Kramer,[5] Jonathan S. Weissman,[1,2,3,*] and Bernd Bukau[5,*]

[1] Howard Hughes Medical Institute

[2] Department of Cellular and Molecular Pharmacology

[3] California Institute for Quantitative Biosciences

[4] Department of Microbiology and Immunology University of California, San Francisco, San Francisco, CA 94158, USA

[5] Center for Molecular Biology of the University of Heidelberg, German Cancer Research Center, DKFZ-ZMBH Alliance, D-69120 Heidelberg, Germany

[6] These authors contributed equally to this work

* Correspondence: weissman@cmp.ucsf.edu (J.S.W.), bukau@zmbh.uni-heidelberg.de (B.B.) DOI 10.1016/j.cell.2011.10.044

SUMMARY

As nascent polypeptides exit ribosomes, they are engaged by a series of processing, targeting, and folding factors. Here, we present a selective ribosome profiling strategy that enables global monitoring of when these factors engage polypeptides in the complex cellular environment. Studies of the *Escherichia coli* chaperone trigger factor (TF) reveal that, though TF can interact with many polypeptides, β-barrel outer membrane proteins are the most prominent substrates. Loss of TF leads to broad outer membrane defects and premature, cotranslational protein translocation. Whereas in vitro studies suggested that TF is pre-bound to ribosomes waiting for polypeptides to emerge from the exit channel, we find that in vivo TF engages ribosomes only after ~100 amino acids are translated. Moreover, excess TF interferes with cotranslational removal of the N-terminal formyl methionine. Our studies support a triaging model in which proper protein biogenesis relies on the fine-tuned, sequential engagement of processing, targeting, and folding factors.

INTRODUCTION

Cotranslational events play a critical role in determining the fate of polypeptides. Indeed, as soon as a nascent chain emerges from the ribosomal exit tunnel, it is acted upon by a series of processing enzymes, targeting factors, and molecular chaperones (Kramer et al., 2009). The ribosome serves as a platform for the regulated association of these various factors. Yet, we have only a limited understanding of the spatial and kinetic coordination of these events.

In bacteria, the exit tunnel of the large ribosomal subunit can accommodate an extended peptide of ~30 amino acids (Ban et al., 2000). Shortly after the peptide exits this tunnel, the formyl group of the N-terminal formylmethionine is removed by a ribosome-bound peptide deformylase (PDF) (Bingel-Erlenmeyer et al., 2008), after which the methionine can be cleaved by methionine aminopeptidase (MAP) (Ball and Kaesberg, 1973). In addition, many nascent polypeptides interact with the ribosome-associated chaperone trigger factor (TF), which is thought to assist in cotranslational folding. Alternatively, the signal recognition particle (SRP) or the ATPase SecA can interact with nascent chains harboring an N-terminal signal sequence in order to target them for translocation across the cytoplasmic membrane (Huber et al., 2011; Ullers et al., 2003). The chaperone SecB also associates with nascent secretion substrates (Randall and Hardy, 2002).

Ribosome-associated chaperones play critical roles in both prokaryotes (Kramer et al., 2009) and eukaryotes (Albanèse et al., 2006; Hundley et al., 2005). Of these, TF is the best characterized in terms of the molecular details of its action (Hoffmann et al., 2010). The ability of TF to promote folding of newly synthesized proteins depends on its association with ribosomal protein L23, which is situated on the surface of the ribosome

near the polypeptide exit channel (Kramer et al., 2002). The ribosome binding activity of TF has been extensively characterized in vitro. Although TF binds to nontranslating ribosomes with a K_D of ~1 mM (Patzelt et al., 2002) and with a mean residence time of 10–15 s (Kaiser et al., 2006), the presence of nascent substrates can increase this affinity up to 30-fold (Rutkowska et al., 2008). In addition, structural analyses of TF in complex with ribosomes suggest that TF forms a protective dome over the tunnel exit (Ferbitz et al., 2004) that could shield nascent chains from degradation by proteases (Hoffmann et al., 2006; Tomic et al., 2006) or improve the efficiency of folding by reducing the speed of folding (Agashe et al., 2004).

By contrast, many aspects of the mechanism of action of TF in vivo are unknown. For example, how TF aids in the folding of proteins remains unresolved. Likewise, it is unclear whether TF interacts with all nascent chains or only a specific subset, and, although TF can interact with relatively short nascent chains in vitro (Merz et al., 2008), it is unknown when TF begins to associate with them in vivo. Furthermore, the interplay of TF with other chaperones, targeting factors, and enzymes remains unclear. Finally, despite extensive studies, the phenotypic cost to cells lacking TF has not been apparent unless combined with the loss of the DnaK chaperone (Deuerling et al., 1999; Teter et al., 1999).

To enable the systematic and quantitative analysis of proteins in prokaryotes, we have developed a strategy for monitoring bacterial translation through ribosome profiling (deep sequencing of ribosome protected mRNA fragments) (Ingolia et al., 2009). Furthermore, by combining ribosome profiling with a procedure to affinity purify ribosomes whose nascent chains are bound by TF, we quantitatively defined when TF engages its substrates. Analysis of these data revealed several fundamental features of TF action, including a role for TF in the biogenesis of β-barrel outer membrane proteins

(OMPs). Additionally, we found that in contrast to in vitro studies, full recruitment of TF is delayed until the peptide is ~100 amino acids in length, providing a protected window during which other processing and targeting factors have preferential access to the nascent chain. More generally, the approach developed here enables the comprehensive and quantitative analysis of cotranslationally acting factors involved in the maturation and folding of newly synthesized polypeptides.

RESULTS AND DISCUSSION

Ribosome Profiling in Escherichia coli

Dramatic advances in DNA sequencing technology (Bentley et al., 2008) have made it possible to sequence bacterial genomes rapidly and at low costs. This has led to an enormous increase in our understanding of the genetic diversity of the prokaryotic world. However, our ability to systematically identify the proteins encoded within these genomes or monitor their rates of production has lagged far behind. Eukaryotic ribosome profiling experiments (Guo et al., 2010; Ingolia et al., 2009) have provided the means to (1) experimentally define open reading frames (ORFs) in an unbiased manner including those that play a regulatory role in translation (or are too small to be identified by other approaches) rather than leading to production of stable proteins; (2) comprehensively evaluate protein production rates for each gene under different environmental conditions; and (3) measure the variability of rates in translation within genes that arise from ribosome pausing at specific positions along the message. We sought to extend this approach to prokaryotes to enable the analysis of both translation and cotranslational processes that promote the maturation of nascent polypeptides. Although we focused on *E. coli*, our method provides a general tool for decoding proteomes and monitoring rates of protein production in other bacteria.

Development of Bacterial Ribosome Profiling

Ribosome profiling requires four distinct steps: (1) generation of cell extracts, in which ribosomes have been halted along the mRNA that they are translating; (2) treatment of polysomes with nuclease to remove regions of the message not protected by the ribosome; (3) conversion of these RNA fragments into double stranded DNA copies; and (4) analysis of these fragments by high-throughput sequencing.

We developed two alternative approaches to capture the cellular state of translation in *E. coli*. For the first, we pretreated exponentially growing cells with chloramphenicol to arrest translating ribosomes. For the second, we collected the cells by fast filtration of the culture. For each case, cells were rapidly frozen in liquid nitrogen and lysed in a frozen state, preventing continued elongation during sample preparation. Both approaches allowed extraction of intact polysomes (Figure 1A), although modest differences in polysome profiles were seen between them. Although rapid filtration is essential for robust analysis of ribosome pausing (see below), chloramphenicol pretreatment is especially useful in cases where rapid recovery of cells is difficult.

After digestion with micrococcal nuclease (MNase), ribosome-protected mRNA footprints were isolated through the collection of monosomes either using a sucrose gradient or by pelleting them through a sucrose cushion. Protected mRNA regions that are derived from other ribosomal complexes (such as disomes) can be distinguished from monosomal footprints based on the size of the protected fragments using PAGE purification. Finally, we converted RNA fragments into a sequenceable DNA library using a previously described method (Ingolia, 2010), except that 3' ends were ligated with a defined linker rather than being polyadenylated. Following conversion, each footprint was identified by deep sequencing and mapped to its genomic position.

Meta-Gene Analysis

Focusing on the top ~2,000 highly expressed genes (out of 4,084 annotated), we analyzed the average ribosome density across these ORFs using cells harvested by rapid filtration. A strong peak was seen over the initiation codon whose density was ~5.5-fold greater than those within the body of the message (Figure 1B, left). A less pronounced peak (~2-fold greater density) was observed over the termination codon (Figure 1B, right). The elevated ribosome density at the beginning and end of coding

sequences presumably reflects the slower kinetics of translation initiation and termination when compared with the average rate of elongation. There was also a modest (~1.3-fold) excess in density over the first 50–100 codons. This is similar in span but of much smaller magnitude to the ~3-fold excess density seen at the 5' end of yeast messages (Ingolia et al., 2009).

Examination of the ribosome occupancy profile of individual genes revealed that the density of ribosome footprints varies substantially across individual messages (Figure 1C), resulting from local differences in the rate of elongation as the ribosome moves down a message. For example, *dnaK* had a median read density of 7.4 reads per million, yet five peaks exceeded this median by more than 10-fold and most likely represent prominent ribosome pausing sites (Figure 1C). This observed variability was highly reproducible ($[R^2] = 0.92$ for *dnaK*) and thus likely represents an intrinsic feature of the translation of individual messages. Ribosome pausing regulates the synthesis (Morris and Geballe, 2000), folding (Zhang et al., 2009), and localization of certain proteins (Mariappan et al., 2010). However, the difficulties in identifying pause sites have limited previous analyses to a small number of examples. Our data provide a critical resource for understanding the causes and biological roles of such pauses.

Defining Open Reading Frames

Ribosome profiling provides a direct readout of the regions being translated along any mRNA, allowing the experimental definition of protein boundaries and thus the identification of novel ORFs. Although the *E. coli* genome has been extensively annotated, we identified a number of short ORFs, including a well-expressed 55 residue protein (Figure 1D) and an upstream uORF with a near cognate (UUG) initiation codon preceding *corA* (Figure 1E). uORFs can regulate the expression of downstream genes in the same operon (Tenson and Ehrenberg, 2002), but their identification has been

challenging. Thus, ribosome profiling provides a general tool for identifying and monitoring production of these species under many environmental conditions independent of their size or stability.

Global Analysis of Gene Expression

Ribosome profiling provides a high precision tool for monitoring translation rates as evidenced by density of ribosomes on messages. Under optimal growth conditions (Luria broth, mid-log phase, 37°C), ~75% of known ORFs were quantifiable. Such measures are highly reproducible ($[R^2] = 0.99$), with rates of translation spanning five orders in magnitude (Figure 1F). This measure of protein expression is expected to be a far better predictor of protein levels than measures of mRNA levels as it captures both transcriptional and translational control (Ingolia et al., 2009). This point is illustrated by analyzing the translation rates of polycistronic messages. Despite being encoded on the same mRNA, the expression levels of genes in the same operon are only modestly correlated with one another (Figure 1G). This finding argues that translational control plays an important role in determining the overall rate of protein production in *E. coli*.

Investigation of TF-Nascent Chain Interactions by Selective Ribosome Profiling

We next sought to extend our technique to selectively profile ribosomes by enriching for ribosomes bound by factors that act on nascent chains. In general, selective ribosome profiling depends on the efficient enrichment of a well-defined population of ribosomes (Figure 2A). Here, we focused on monitoring monosomes that were engaged by TF, predominantly through its association with the nascent chain.

To facilitate the purification of TF-bound ribosomes, we fused a tag consisting of a TEV protease-cleavable AviTag, which is biotinylated by an endogenous biotin ligase, to the C-terminus of the protein. The tagged TF appears to be fully functional both in

vitro and in vivo. The tag neither altered the affinity of TF for the ribosome nor interfered with TF's ability to aid refolding of chemically denatured GAPDH (data not shown). Furthermore, expression of tagged TF at wild-type levels complemented the synthetic lethal phenotype seen for the Δ*tig* Δ*dnaK* double knockout (Figures S1A and S1B available online) and the chemical sensitivities of Δ*tig* cells (see below) (Figures S5A and S5B).

To stabilize the transient association of TF with ribosome-nascent chain complexes (TF-RNCs), we crosslinked TF to nascent polypeptides using the thiol-cleavable crosslinker DSP (dithiobis succinimidyl propionate), which reacts with primary amines in lysine side chains and N-termini. In order to capture physiologically relevant substrate interactions, frozen lysates were directly thawed in the presence of DSP. Lysates were subsequently treated with MNase to generate monosomes prior to affinity purification in order to avoid the co-purification of unbound ribosomes tethered through the polysomal mRNAs. The ribosomal fraction was separated from uncrosslinked TF by ultracentrifugation through a high salt sucrose cushion, followed by affinity purification and elution of TF-RNCs.

Our analysis of crosslinking products demonstrates that we specifically enriched for ribosomes whose nascent chains were engaged by TF. Robust recovery of ribosomes strictly depended on DSP crosslinking (Figure 2B, lanes 4 and 8), the presence of an AviTag (Figure 2B, lanes 1 and 5), and a TEV protease cleavage site (Figure 2B, lanes 2 and 6). Importantly, ribosome recovery also depended on the ability of TF to bind ribosomes (Figure S1C), indicating that TF is unable to directly engage nascent chains without docking to L23. Crosslinking of TF by DSP gave rise to products of diverse molecular weight, representing nascent chains of various lengths (Figure 2B, lane 3, i and ii). However, we observed only negligible crosslinking of TF to L23, since

L23 migrated almost exclusively as a single band under both nonreducing (Figure 2B, lane 3, iii) and reducing conditions (Figure 2B, lane 7, iii). Likewise, we did not observe significant levels of crosslinking between TF and ribosomal proteins L24 or L29 (data not shown), which were suggested to come in close proximity to ribosome-bound TF (Baram et al., 2005; Schlünzen et al., 2005). These observations argue that the purified TF-RNCs were captured predominantly on the basis of the interaction of TF with nascent chains.

Features of TF Engagement to Nascent Chains

We next compared the density of ribosome footprints across individual genes for the affinity purified TF-RNCs to the total pool of ribosomes. The ratio of these values provides a position specific measure of the propensity of TF to engage nascent chains. We performed a meta-gene analysis to determine the average enrichment efficiency as a function of polypeptide length. Since DSP specifically crosslinks TF to nascent chains, the minimal length at which nascent chains engage TF is expected to exceed the 30 amino acids needed to traverse the ribosomal exit tunnel. Indeed, ribosomes within this region were poorly captured by affinity purification (Figure 3A). However, this N-terminal depletion extended well beyond the minimal length needed for the nascent chain to emerge from the ribosome, indicating that effective TF binding requires substantial extension of the polypeptide outside of the tunnel exit.

By examining individual profiles, we found that enrichment efficiency was low near the N-terminus and typically rose sharply thereafter (Figure 3B). The position of this rise (indicative of the first TF binding) varied between different genes. Thus, to identify the initial point at which TF engaged each polypeptide, we measured the position at which each profile first crossed an empirically derived threshold. This threshold was chosen to be well above background but still able to capture these early binding events

(Figure 3C and Extended Experimental Procedures). The median length at which TF first engaged a polypeptide was 112 amino acids, with half of all nascent chains being bound within ±20 amino acids of this position. TF was depleted at the beginning of translation in virtually all nascent chains (Figures 3B and 3C).

We also observed from the meta-gene analysis that the intensity of TF engagement leveled off after ~135 residues (Figure 3A). If multiple TF molecules were bound per nascent chain, the likelihood of purifying crosslinked TF-RNCs should increase in proportion to the length of the nascent chain. Yet, this was not the case, suggesting that nascent chains are generally not engaged by multiple TF molecules. Consistent with this view, we observed varying levels of TF engagement for each polypeptide (Figure 3B), with periods of TF binding interrupted by regions with little detection of TF association over background, suggesting that TF cycles on and off the nascent chain during synthesis.

TF Recruitment by the Ribosome Occurs Concurrently with Nascent Chain Binding

The depletion of TF in the beginning of translation could arise either because TF is not bound to the ribosome or because it is present on the ribosome but not engaged with the nascent chain. To discriminate between these possibilities, we used the crosslinker EDC (1-ethyl-3-[3-dimethylaminopropyl] carbodiimide), which couples carboxyl groups to primary amines. In contrast to DSP, EDC covalently linked TF not only to nascent chains, but also to the ribosome at L23 (Figures 4A and S2). This ability of EDC, unlike DSP, presumably reflects the availability of carboxyl groups in L23 that are in close proximity to TF. Remarkably, the TF enrichment efficiency was highly similar between DSP and EDC at both the meta-gene (Figure 4B) and single gene (Figures 4C and 4D) levels, suggesting that robust ribosome binding occurs concurrently with

nascent chain engagement.

Nascent Chain N-Termini Are Resistant to Surveillance by TF as They Emerge from the Ribosomal Exit Tunnel

In principle, the observed delay in TF recruitment to ribosomes until ~100 amino acids have been synthesized could result from a general paucity of binding sites robustly recognized by TF. Alternatively, the observed depletion could be an intrinsic feature of translation in vivo that disfavors TF recruitment to the N-termini of nascent chains even if TF recognition sites are present. To discriminate between these possibilities, we used the selective ribosome profiling approach to monitor TF engagement in cells expressing variants of a TF substrate altered at their N-termini. Specifically, we constructed a series of OmpF variants in which 48 or 96 residues had been truncated from the N-terminus, 50 residues derived from human myoglobin had been added following the signal sequence, or charged residues (N5D, V9E, A13D, and V16E) had been introduced in the signal sequence. These variants were expressed from a plasmid at levels similar (within a factor of two) to endogenous *ompF* expression (Figure S3). Selective ribosome profiling experiments were then performed to determine when TF engaged the different OmpF variants.

Our results establish that the distance from the N-terminus is the critical determinant of TF recruitment. For each of the truncations and the insertion mutant, there was a complete lack of TF recruitment prior to translation of 50 amino acids, and the synthesis of at least ~100 amino acids was required for the full engagement of TF, even though TF binding sites were present earlier (Figure 5A, i and ii). Following the initial binding event, the pattern of TF engagement along the nascent chain mirrored the binding pattern seen for wild-type OmpF. The signal sequence mutant had no discernible effect on TF binding (Figure 5A, iii), further emphasizing that initial TF engagement

depends on the position along the nascent chain, rather than sequence composition of the residues near the N-terminus.

The delay in TF recruitment to the ribosome—until well after the polypeptide emerges from the exit tunnel—contrasts with the current view, mainly drawn from in vitro data, that TF is pre-bound to the ribosome and waits for the nascent chain to emerge (Patzelt et al., 2002). It is presently unclear what prevents TF from associating with shorter polypeptides in vivo. Although TF is thought to be in excess of ribosomes (Patzelt et al., 2002), the fraction of TF molecules available for nascent chain and ribosome binding at steady-state conditions is not known. For example, TF has recently been suggested to have an additional ribosome-independent function in the assembly of oligomeric complexes (Martinez-Hackert and Hendrickson, 2009). As a result, fewer TF molecules would be available to interact with ribosomes. This could drive TF to associate preferentially with translating ribosomes exposing longer nascent chains. Indeed, such RNCs have been demonstrated in vitro to exhibit higher association rates for TF binding (Rutkowska et al., 2008).

Regardless of the mechanism, this delayed association of TF to RNCs may provide a window for other ribosome-associated nascent chain interacting factors, such as PDF and MAP, to act on the emerging polypeptide. To investigate the interplay of TF with PDF and MAP, we developed an in vitro assay for examining the action of these processing enzymes. We monitored the synthesis of the TF model substrate barnase (which has only one methionine residue at the initiation codon) by following ^{35}S-methionine incorporation in a translation-competent Δ*tig* extract devoid of PDF and MAP activity. In the absence of both enzymes, we detected a pronounced band corresponding to full-length, nonprocessed barnase (Figure 5B, lane 1). In their presence, however, the radioactive signal dramatically decreased (Figure 5B, lane 2), indicating that the N-

terminal methionine was both deformylated and cleaved. Addition of excess TF prior to translation initiation (Figure 5B, lane 3), but not of the TF mutant impaired in ribosome binding (Figure 5B, lane 4) partially restored the radioactive signal, indicating that ribosome-bound TF interferes with N-terminal processing. Similarly, even a modest (~2-fold) overexpression of TF resulted in increased sensitivity to the PDF inhibitor actinonin (Figure 5C). These results suggest that TF can be driven to engage the N-termini of nascent chains, but that premature engagement of TF interferes with the removal of the N-terminal fMet residue from nascent chains, i.e., the essential N-terminal processing carried out by PDF and MAP. More broadly, our data suggest a model in which initial binding of TF to RNCs is determined by the length of the polypeptide, providing access for other factors. After the initial engagement event, TF can repeatedly bind to and release from the nascent chain and may stay associated with it even after translation has terminated (Figure 5D).

TF Interacts with Cytoplasmic Proteins but Shows Strong Preference for Outer Membrane β-Barrel Proteins

To characterize the substrate specificity of TF, we determined the overall enrichment efficiency for each gene—defined as the sum of the enriched footprint density divided by the sum of the total footprint density (excluding the N-terminal region not engaged by TF). This analysis revealed an apparent bimodal distribution, with a major subset comprised of nascent chains robustly interacting with TF and a minor subset showing modest engagement to the chaperone (Figure 6A). Whereas most nascent chains actively engaged by TF were localized to the cytoplasm ($p = 5.2 \times 10^{-16}$, rank sum test), those poorly engaged by the chaperone were generally targeted for the inner membrane ($p = 3.2 \times 10^{-44}$) (Figure 6B). This division of the TF interactome along cellular localization is consistent with the view that SRP outcompetes TF for binding to

nascent inner membrane proteins (Eisner et al., 2006; Ullers et al., 2003, 2006).

Strikingly, outer membrane β-barrel proteins were among the strongest TF interactors (Figure 6B). For example, five of the best characterized OMPs (LamB, LptD, OmpA, OmpC, and OmpF) were among the top 25 most strongly enriched polypeptides (out of the ~2,000 proteins examined). In addition, LamB, OmpA, OmpC, and OmpF were at least an order of magnitude more strongly expressed than the other proteins that were highly enriched in the TF pull downs. Thus they accounted for the large majority of polypeptides fluxing through TF among this enriched group. Interestingly, TF was originally identified on the basis of its ability to promote insertion of chemically denatured pro-OmpA into membrane vesicles (Crooke and Wickner, 1987), although the physiological significance of this was unclear (Guthrie and Wickner, 1990).

Loss of TF Function Mimics the Loss of Outer Membrane Chaperones

The enrichment of OMPs among the cotranslational substrates of TF suggested that TF could play a role in OMP biogenesis. Defects in the biogenesis of outer membrane β-barrel proteins (including porins) often disrupt outer membrane integrity, leading to increased sensitivity to SDS/EDTA and vancomycin (Hagan et al., 2011). In support of the notion that TF plays a role in OMP biogenesis, Δ*tig* cells displayed increased sensitivity to both SDS/EDTA (Figures 6C and S4) and vancomycin (Figures 6D and S4) in a manner that was rescued by wild-type TF or tagged TF expressed from plasmid (Figures S5A and S5B). Additionally, the activity of σ^E, which controls the envelope stress response, decreased by ~2-fold (Figure S6). As σ^E monitors the protein flux and integrity of folding in the outer membrane, the lowered activity of σ^E could reflect altered delivery of OMPs to the periplasm (Mecsas et al., 1993).

To further examine the phenotypic consequences of the loss of TF function, we

compared the chemical sensitivities of Δ*tig* cells to ~4,000 deletion strains that were exposed to more than 300 conditions in a large-scale chemical genetic screen (Nichols et al., 2011). Because mutations in functionally related genes have closely related chemical sensitivities (Hillenmeyer et al., 2008), examining the correlation across these sensitivities can reveal relationships among either unknown or seemingly unrelated genes. The chemical sensitivities of Δ*tig* cells correlate highly with those carrying mutations in *bamA* ($p = 1.5 \times 10^{-11}$), *bamB* ($p = 8.0 \times 10^{-12}$), *bamE* ($p = 8.2 \times 10^{-11}$), *surA* ($p = 5.8 \times 10^{-11}$), and *yfgC* ($p = 6.0 \times 10^{-15}$) (Figure 6E). *bamA*, *bamB*, and *bamE* encode for three of five components of the β-barrel assembly machinery, which together with SurA (a periplasmic chaperone) mediates the insertion of β-barrel proteins into the outer membrane (Hagan et al., 2011), strongly implicating a role for TF in OMP biogenesis. *yfgC* encodes for a predicted periplasmic peptidase and although its function is unknown, the chemical genetic data suggests that it also plays a role in OMP biogenesis. Indeed, when we queried the chemical sensitivities of Δ*yfgC* cells against all other strains, we found that these correlations were among the highest with components of the Bam complex and TF (Figure 6F).

The phenotypic link between SurA/Bam(s) and TF is particularly remarkable in light of the structural similarity between SurA and TF. SurA and TF have no apparent sequence similarity, yet their chaperone domains possess the same fold (Merz et al., 2006). Moreover, the chaperone domain of both proteins is followed by *cis*/trans peptidyl prolyl isomerase domain(s). Although the functional significance of these observations was not evident at the time, we now suggest the possibility that there was a primordial chaperone for β-barrel proteins, which underwent a gene duplication event, such that one copy gained (or retained) a ribosome-binding domain (TF), whereas the other (SurA) gained an N-terminal signal sequence, targeting it to the periplasm. This could then give

rise to a pathway capable of chaperoning β-barrel proteins from their synthesis by the ribosome to their insertion into the outer membrane.

To provide quantitative data on the outer membrane defect caused by TF deletion, we analyzed the protein content of the outer membrane of wild-type and Δ*tig* cells using SILAC (stable isotope labeling with amino acids in cell culture) combined with mass spectrometric analysis. We found that ~60% of all detected proteins in the outer membrane fraction showed a significant decrease in their steady-state levels in Δ*tig* cells compared with wild-type cells, whereas almost no proteins were found in higher amounts upon TF deletion (Figure 7A). OmpA and components of the Bam complex were among the most prominently disenriched proteins. Levels of OmpC and OmpF remained unaltered, which may result from compensatory mechanisms that are known to regulate their transcription so as to maintain proper OmpC/OmpF levels in the outer membrane. Interestingly, when we also analyzed the soluble fraction of proteins by SILAC, we found that SecA is one of the most strongly induced proteins in the Δ*tig* mutant (increased by ~40%, $p = 2.6 \times 10^{-4}$), suggesting that Δ*tig* cells have a mild translocation defect (Riggs et al., 1988).

The Impact of Loss of TF on OMP Translocation

Taken together, the above observations indicate that an important function of TF is to chaperone β-barrel OMPs and/or modulate their export to the periplasm by the translocation system. Consistent with previous studies (Lee and Bernstein, 2002; Ullers et al., 2007), we observed that Δ*tig* mutants accumulated less full-length precursors in pulse-labeling experiments for all exported proteins we examined (LamB, MBP, OmpA, and OmpF), suggesting that signal sequence processing (and therefore protein translocation) is more cotranslational in strains lacking TF. To test for this directly, we used an established two-dimensional gel assay (Josefsson and Randall, 1981) to

determine if the timing of translocation with respect to nascent polypeptide length was affected in a Δ*tig* mutant. ^{35}S-labeled LamB was immunoprecipitated and resolved by size using SDS-PAGE. The protein was then subjected to partial proteolysis in gel with a site-specific protease (V8), with the proteolytic fragments separated by size in a second dimension. For wild-type cells, the prominent C-terminally derived fragments converged on the pro-LamB form, implying that a substantial fraction of the polypeptides were exported only after completion of protein synthesis (Figures 7B and S7B for a similar analysis of MBP in which the N-terminal fragments are more prominent). However, for Δ*tig* cells, the C-terminal fragments converged on the mature form, indicating that there was significantly more cotranslational export of LamB. This switch in translocation mechanism helps explain the suppression of the translocation defect of a *secB* mutant by mutations in the *tig* gene (Ullers et al., 2007). In addition, the mechanism of translocation can affect the folding of exported proteins in the periplasm (Kadokura and Beckwith, 2009), which can partially explain the apparent OMP defect of Δ*tig* mutants. More generally, our findings underscore the challenge in attempts to define the function of components in highly redundant systems by following the phenotypes resulting from the loss of these components.

Perspective

Here, we present a strategy for the quantitative analysis of translation in bacteria using ribosome profiling. This technique provides a critical tool for decoding unknown bacterial proteomes, quantitatively monitoring translation rates, and exploring the various mechanisms for regulating translation. Additionally, we present an approach to selectively profile ribosomes, which enabled us to query the substrates of the cotranslationally acting chaperone TF. Our studies revealed that recruitment of TF is

delayed until well after the polypeptide has emerged from the ribosome exit channel. This delay is likely to be critical in allowing other factors to engage nascent chains. In support of this notion, excess TF prevents N-terminal processing both in vitro and in vivo. Selective ribosome profiling of other factors should help decipher the logic underlying the coordinated action of the various ribosome-associated processing enzymes, targeting factors and molecular chaperones that ensure the efficient biogenesis of proteins in vivo.

The value of selective ribosome profiling is also illustrated by the identification of OMPs as critical targets of TF. Defining the function of TF has been challenging because of its redundancy with other chaperone systems, which masks the phenotypic consequence of loss of TF. By contrast, our approach can monitor the natural flux of TF substrates in unperturbed cells, which revealed that abundantly expressed OMPs were consistently among the most prominent substrates of TF. This was complemented by analysis of a comprehensive chemical genetic screen, which showed that the chemical sensitivities of the loss of TF closely resembled that seen with the loss of the OMP chaperone machinery, as well as global mass spectrometry analysis. We anticipate that this combination of quantitative phenotypic loss of function analysis and high resolution monitoring of chaperone action in unperturbed cells will be key to elucidating the in vivo function of chaperone networks.

ADDITIONAL INFORMATION

Experimental Procedures

General Ribosome Profiling

Bacterial cells were grown in LB media at 37°C to an OD_{600} of 0.4–0.5. Cells were harvested either by pretreatment with chloramphenicol to a final concentration of 100 mg/ml or by rapid filtration. Collected cells were flash frozen in liquid nitrogen and cryogenically pulverized by mixer milling (Retsch). Pulverized cells were thawed and clarified by centrifugation. Resulting lysates were digested with MNase, quenched with EGTA, and resolved by sucrose density gradients. Ribosome-protected mRNA footprints were processed as previously described (Ingolia, 2010) and deep sequenced by Illumina GA II or HiSeq2000.

Selective Ribosome Profiling

Bacterial cells were collected and pulverized as with the general approach, but cells were thawed directly in the presence of 2.5 mM DSP or 20 mM EDC (pH 5.8) (Pierce). To quench the crosslinking reactions, lysates were brought up to 100 mM Tris (pH 8.0) for DSP or 100 mM Tris (pH 8.0), 250 mM glycine, and 4 mM $NaHCO_3$ for EDC. Crosslinked lysates were digested with MNase and resolved by sucrose density gradients or cushions. Ribosome pellets derived from sucrose cushions were resuspended in buffer containing 50 mM Tris (pH 7.0), 200 mM NaCl, 10 mM $MgCl_2$, 1 mM chloramphenicol, 1 mM PMSF, 0.4% Triton X-100, and 0.1% NP-40 and incubated overnight on ice. The resuspended ribosome mixture was affinity purified with 220 μl of 50% Strep-Tactin Sepharose (IBA, Germany) and thoroughly washed. TF-RNCs were next eluted by TEV protease treatment at room temperature for 1 hr. Typical yields ranged between 50 and 100 μg of RNA as determined by Nanodrop (Thermo Scientific).

Ribosome-protected mRNA fragments were isolated and converted to a cDNA library for identification.

Accession numbers

Sequences for *nadS* and *corL* were deposited in GenBank (http://www.ncbi. nlm.nih.gov/genbank/) with accession numbers JQ045772 and JQ045773. Sequencing data were deposited in the GEO database (http://www.ncbi. nlm.nih.gov/geo/) with accession number GSE33671.

Supplemental Information

Supplemental Information includes Extended Experimental Procedures and seven figures, and can be found with this article online at doi:10.1016/j.cell. 2011.10.044.

Acknowledgements

We thank N. Ingolia for help with ribosome profiling and data analysis; C. Chu for help with sequencing; S. Rouskin for developing a method for ligation; G. Brar and J. Dunn for developing an rRNA subtraction strategy; N. Reifenberger and B. Zachmann-Brand for help with cloning and crosslinking experiments; T. Silhavy for providing antibodies; T. Ruppert and the mass spectrometry facility at the ZMBH for experimental support; Y. Cully for graphical assistance; and members of the Bukau, Gross and Weissman labs for insightful discussions and critical readings of the manuscript. This work was supported by the SFB638 and FOR967 of the Deutsche Forschungsgemeinschaft (B.B. and G.K.); the Peter und Traudl Engelhorn-Stiftung and EMBO (A.H.B.); and the NIH (P01 AG10770) and the Howard Hughes Medical Institute (J.S.W.).

Figure 1

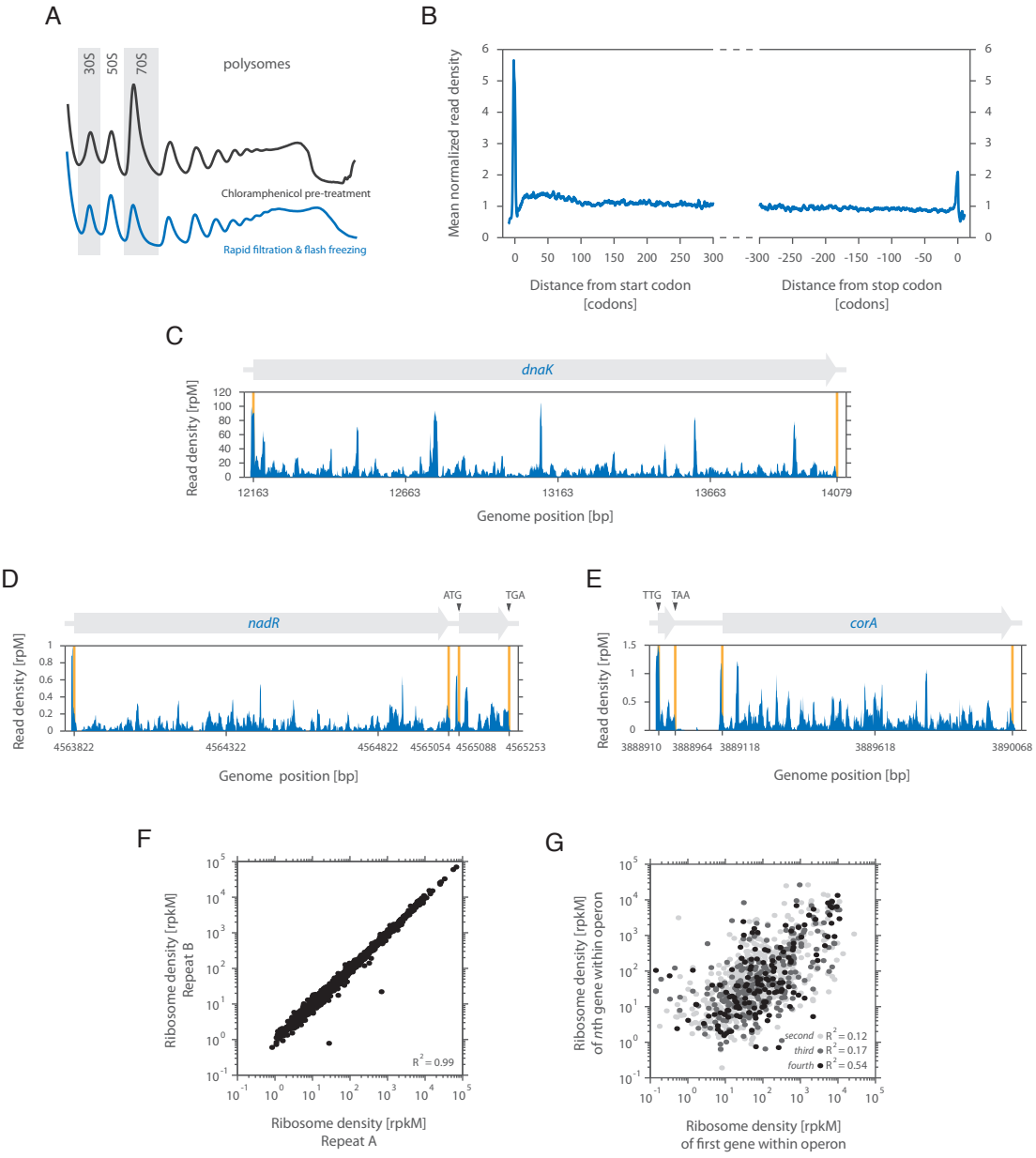

A
30S
50S
70S
polysomes
Chloramphenicol pre-treatment
Rapid filtration & flash freezing
B
Mean normalized read density
Distance from start codon [codons]
Distance from stop codon [codons]
C
dnaK
Read density [rpM]
Genome position [bp]
D
nadR
ATG
TGA
Read density [rpM]
Genome position [bp]
E
corA
TTG
TAA
Read density [rpM]
Genome position [bp]
F
Ribosome density [rpkM] Repeat B
Ribosome density [rpkM] Repeat A
R2 = 0.99
G
Ribosome density [rpkM] of nth gene within operon
Ribosome density [rpkM] of first gene within operon
second R2 = 0.12
third R2 = 0.17
fourth R2 = 0.54

Figure 1

Characterizing Prokaryotic Translation by Ribosome Profiling of Bacterial Cells

(A) Translating ribosomes were extracted from cells (MC4100) either pretreated with chloramphenicol (black trace) or collected by rapid filtration (blue trace). Polysomes were resolved by 10%–55% (w/v) sucrose density gradients. **(B)** Meta-gene analysis of ribosome density as a function of position from fast filtered cells. Genes were aligned from either their start (left panel) or stop (right panel) codon and averaged across them (see Extended Experimental Procedures). **(C)** Ribosome density of *dnaK* as a function of position. The density in reads per million (rpM) was corrected for the total number of reads that aligned to all coding sequences. **(D)** Example of a newly identified canonical ORF *nadS* (GenBank accession number JQ045772). **(E)** Example of a novel ORF *corL* starting at a noncanonical UUG codon (GenBank accession number JQ045773). **(F)** Quantifying gene expression levels by ribosome profiling from fast filtered cells. Ribosome densities of two independent replicates were plotted for comparison. The density in reads per kilobase million (rpkM) is a measure of overall translation along each gene (see Extended Experimental Procedures). **(G)** The ribosome density of the first gene in an operon was compared with the ribosome density of either the second, third, or fourth gene in the same operon as indicated.

Figure 2

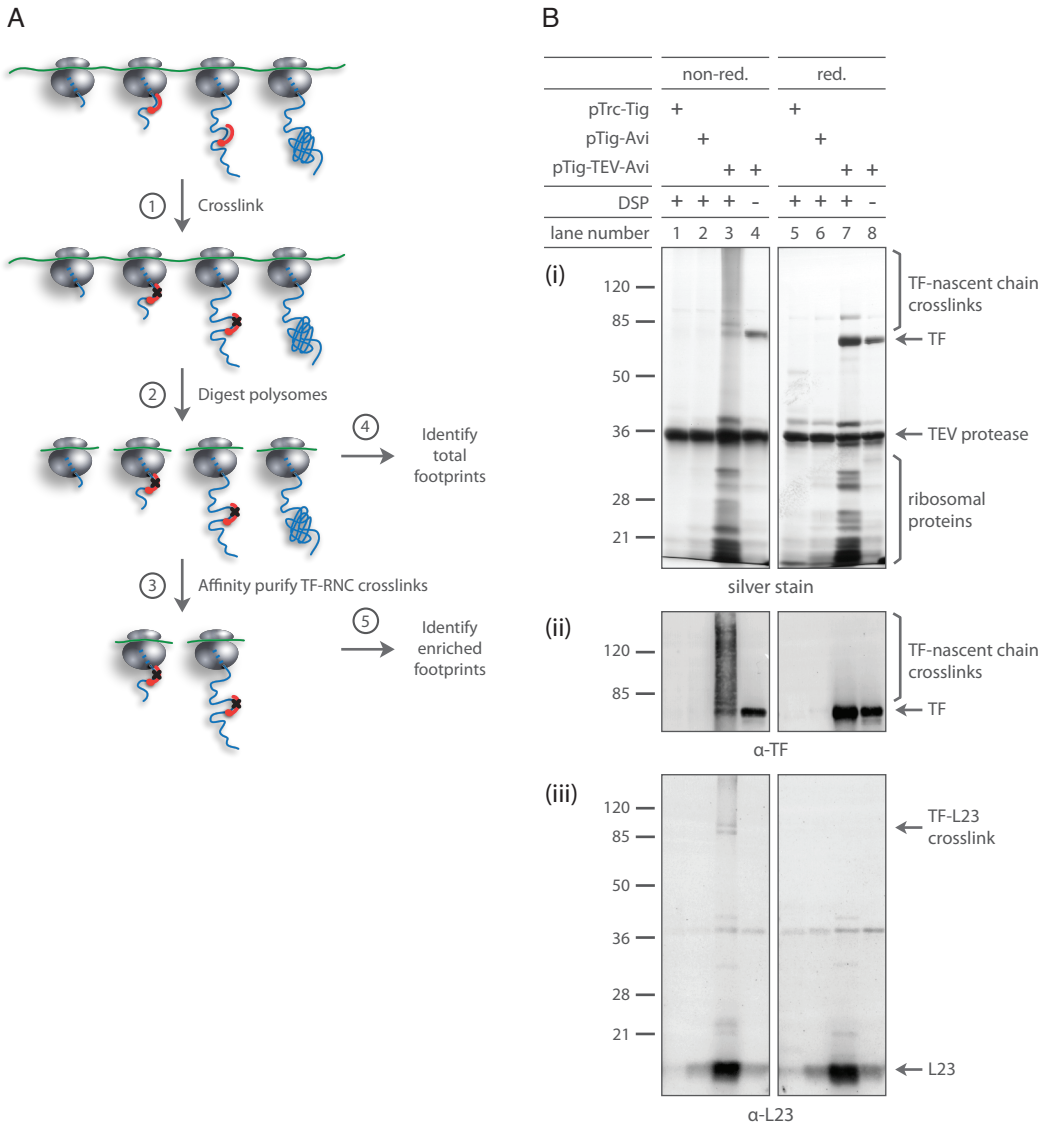

A
1 Crosslink
2 Digest polysomes
3 Affinity purify TF-RNC crosslinks
4 Identify total footprints
5 Identify enriched footprints
B
non-red.
red.
pTrc-Tig
pTig-Avi
pTig-TEV-Avi
DSP
lane number
1 2 3 4 5 6 7 8
(i)
120
85
50
36
28
21
TF-nascent chain crosslinks
TF
TEV protease
ribosomal proteins
silver stain
(ii)
120
85
TF-nascent chain crosslinks
TF
α-TF
(iii)
120
85
50
36
28
21
TF-L23 crosslink
L23
α-L23

Figure 2

TF Crosslinked RNCs Can Be Isolated with High Specificity

(A) Schematic for affinity purifying TF crosslinked RNCs: (1) cells expressing epitope-tagged TF are harvested at mid-log phase, cryogenically lysed, and chemically crosslinked. (2) Polysomes digested with MNase yield footprint-containing monosomes. (3) Digested monosomes are forced through a sucrose cushion, separating free TF molecules from those crosslinked to RNCs. TF crosslinked RNCs are affinity purified and eluted by cleaving TF with TEV protease. (4) mRNA footprint fragments derived from all monosomes and (5) those enriched through affinity purification are cloned into a cDNA library for deep sequencing analysis. **(B)** Gel analysis of DSP crosslinking and affinity purification. Δ*tig::kan* cells expressing specified TF variants were harvested by centrifugation. Following cryogenic lysis, lysates were crosslinked with DSP as they thawed. TF-RNCs were affinity purified and eluted through TEV protease cleavage. Eluates were analyzed under both reducing and nonreducing conditions. Gels were either silver stained (i) or immunoblotted using antisera specific for TF (ii) or L23 (iii).

Figure 3

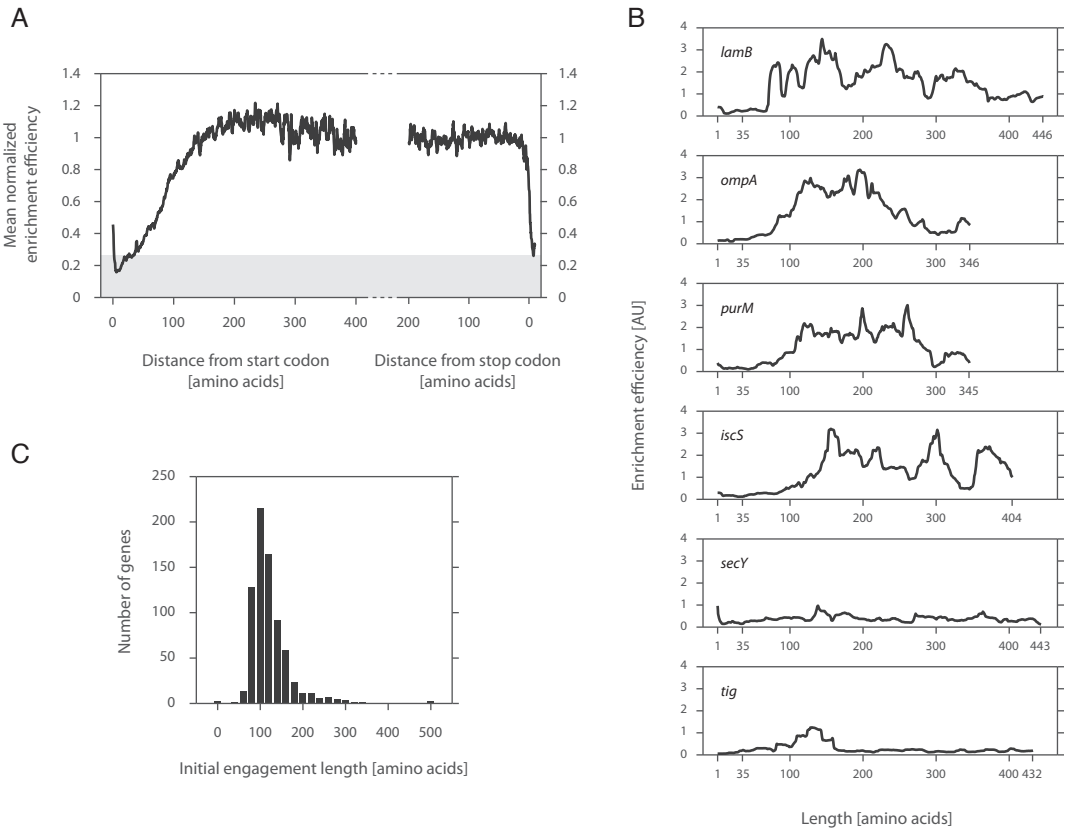

A
Mean normalized enrichment efficiency
Distance from start codon [amino acids]
Distance from stop codon [amino acids]
B
lamB
ompA
purM
iscS
secY
tig
Enrichment efficiency [AU]
Length [amino acids]
C
Number of genes
Initial engagement length [amino acids]

Figure 3

TF Interaction Propensity as a Function of Nascent Chain Length

(A) Meta-gene enrichment efficiency derived as a function of ribosome position. Meta-gene ribosome densities (described in Figure 1B) were each computed for footprints derived from TF enriched RNCs and those from the total monosome pool. Ratios between these profiles were taken along indicated positions. Background signal is shaded in gray, corresponding to the enrichment efficiency at codon 30, a length that should be inaccessible to soluble factors. **(B)** Individual enrichment efficiencies plotted as a function of nascent chain length. Characteristic examples of cytoplasmic (IscS and PurM), inner membrane (SecY), and outer membrane (LamB and OmpF) proteins are shown. **(C)** A histogram of the initial position at which TF engages nascent chains.

Figure 4

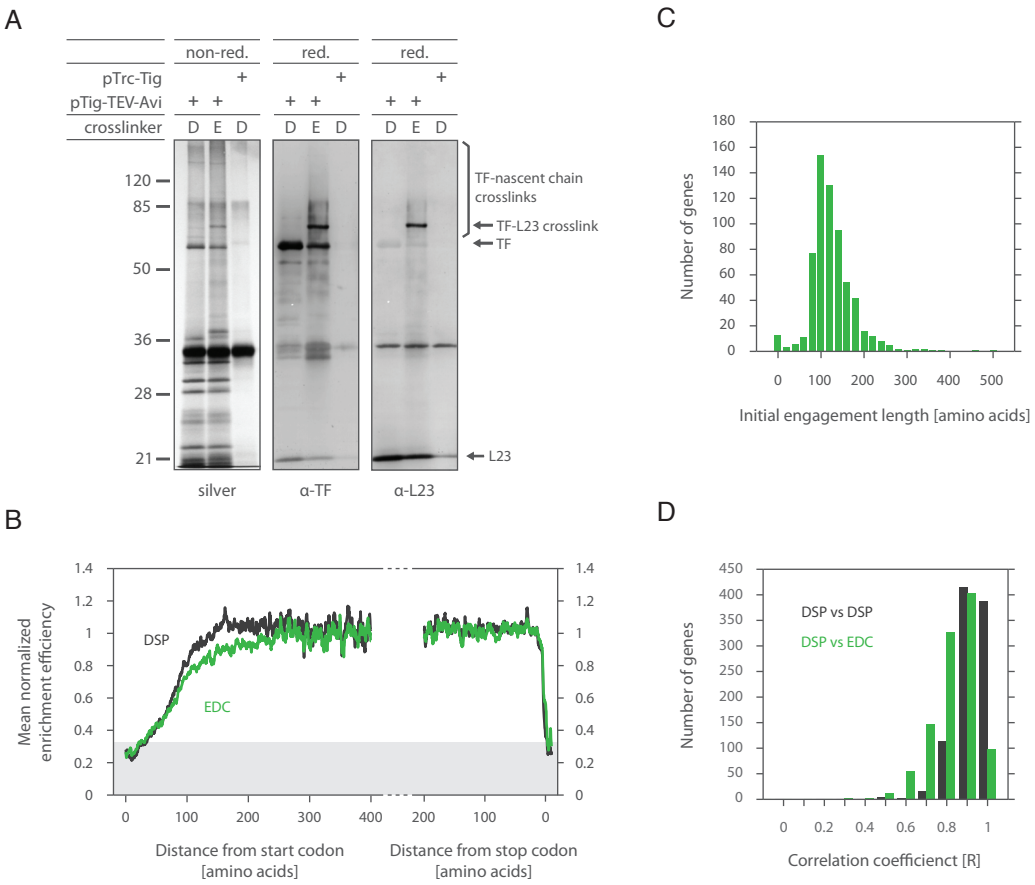

A
non-red.
red.
red.
pTrc-Tig
pTig-TEV-Avi
crosslinker
D E D D E D D E D
120
85
50
36
28
21
TF-nascent chain crosslinks
TF-L23 crosslink
TF
L23
silver
α-TF
α-L23
B
Mean normalized enrichment efficiency
DSP
EDC
Distance from start codon [amino acids]
Distance from stop codon [amino acids]
C
Number of genes
Initial engagement length [amino acids]
D
Number of genes
DSP vs DSP
DSP vs EDC
Correlation coefficienct [R]

Figure 4

Ribosome Recruitment of TF Occurs at the Same Time as Nascent Chain Binding

(A) Gel analysis of DSP and EDC crosslinking and affinity purification. Δ*tig::kan* cells were processed as before (Figure 2), but treated with DSP (D) or EDC (E). Resulting eluates were resolved by SDS-PAGE under both reducing (red.) and nonreducing (non-red.) conditions. EDC crosslinks are irreversible under reducing conditions unlike DSP. Gels from nonreduced samples were silver stained (i), whereas reduced samples were immunoblotted using antisera specific either for TF (ii) or L23 (iii). **(B)** The same test was performed as in Figure 3A, except cells were harvested by rapid filtration followed by fast freezing. Cryogenically pulverized cells were crosslinked with either EDC or DSP. **(C)** The same test was performed as in Figure 3C, except cells were harvested by filtration (as in Figure 4B) and TF-RNCs were stabilized by EDC crosslinking. (D) Gene by gene correlation of TF binding profiles for DSP replicates and DSP compared to EDC.

Figure 5

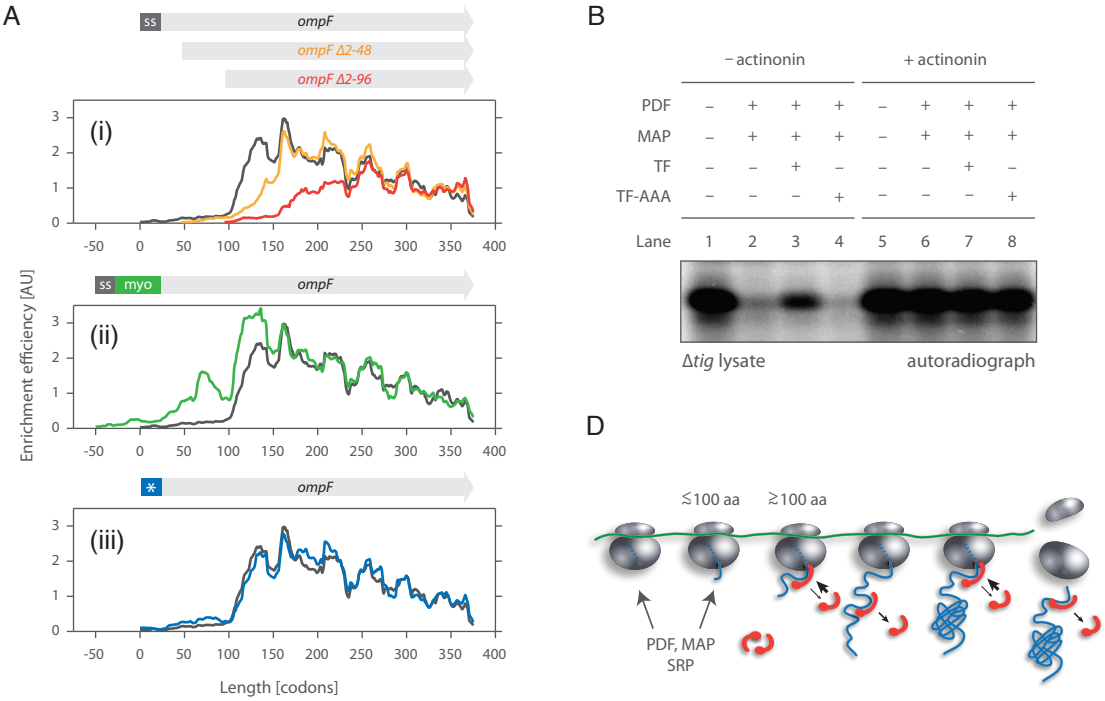

A
ss
ompF
ompF Δ2-48
ompF Δ2-96
(i)
ss myo
ompF
(ii)
*
ompF
(iii)
Enrichment efficiency [AU]
Length [codons]
B
– actinonin
+ actinonin
PDF
MAP
TF
TF-AAA
Lane
1 2 3 4 5 6 7 8
Δtig lysate
autoradiograph
D
≲100 aa
≳100 aa
PDF, MAP
SRP
C
0 μM IPTG
10 μM IPTG
20 μM IPTG
30 μM IPTG
pTrc99
pTrc-Tig
pTrc-TigAAA
0.1 μg/ml actinonin
No drug control

Figure 5

TD Does Not Engage the N-Terminal End of Nascent Chains In Vivo as They Emerge from the Ribosome

(A) Individual enrichment efficiencies of *ompF* variants compared with wild-type *ompF*. **(B)** Cell-free coupled transcription/translation reactions initiated by nonstalled barnase. Five micromolars each of either TF or TF-AAA (deficient in ribosome binding) and 2 μM each of both PDF and MAP were supplemented prior to translation initiation where indicated. Extracts treated with actinonin (lanes 5–8), an inhibitor of PDF, were used to assess overall levels of barnase synthesis. Reactions were quenched by TCA precipitation and visualized using SDS-PAGE and autoradiography. **(C)** TF overexpression can interfere with N-terminal processing. MC4100 Δ*acrA::kan* cells transformed with pTrc99 (empty vector), pTrc-Tig, or pTrc-TigAAA were spotted as 1:10 serial dilutions on LB plates containing 100 μg/ml of ampicillin and indicated concentrations of IPTG and actinonin. Ten micromolars IPTG induces TF from pTrc-Tig near endogenous levels (Kramer et al., 2004); thus, overall expression of TF is increased roughly 2-fold (from both the plasmid and endogenous locus). **(D)** A model for the dynamic binding of TF to ribosomes and nascent chains. Interaction between TF and the ribosome is limited early on translation (i.e., before the nascent chain emerges from the exit tunnel). TF fully engages the ribosome and nascent chain not before ~100 amino acids are translated. After release from the nascent chain, TF can rebind, but each polypeptide is, on average, bound by only one TF molecule at a time. Following translation termination, TF may stay associated with the released polypeptide, guiding further folding steps.

Figure 6

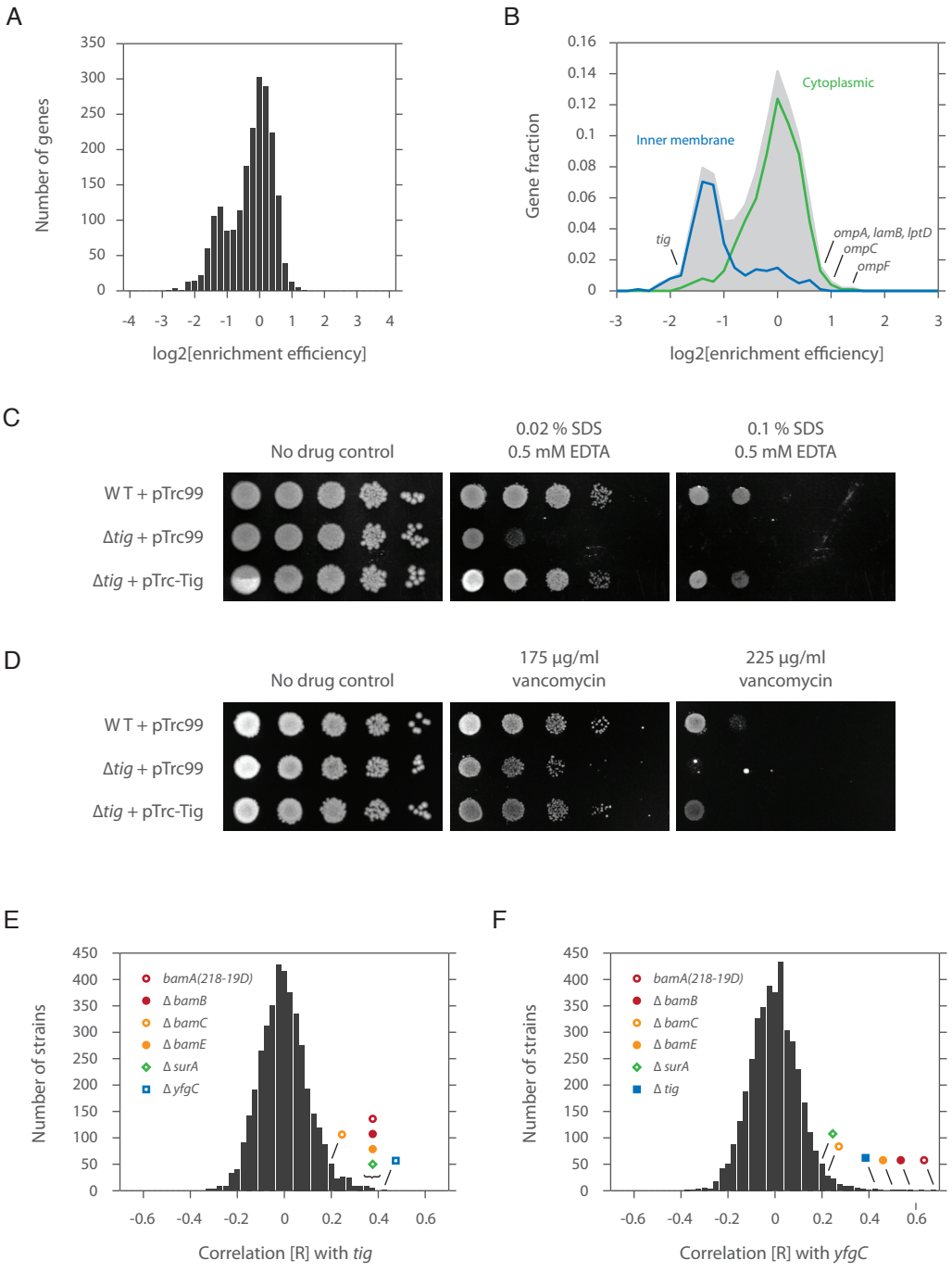

A
Number of genes
log2[enrichment efficiency]
B
Gene fraction
Cytoplasmic
Inner membrane
tig
ompA, lamB, lptD
ompC
ompF
log2[enrichment efficiency]
C
No drug control
0.02 % SDS
0.5 mM EDTA
0.1 % SDS
0.5 mM EDTA
WT + pTrc99
Δtig + pTrc99
Δtig + pTrc-Tig
D
No drug control
175 µg/ml
vancomycin
225 µg/ml
vancomycin
WT + pTrc99
Δtig + pTrc99
Δtig + pTrc-Tig
E
Number of strains
bamA(218-19D)
Δ bamB
Δ bamC
Δ bamE
Δ surA
Δ yfgC
Correlation [R] with tig
F
Number of strains
bamA(218-19D)
Δ bamB
Δ bamC
Δ bamE
Δ surA
Δ tig
Correlation [R] with yfgC

Figure 6

TF Chaperones Outer Membrane β-Barrel Protein Biosynthesis

(A) A histogram of the overall enrichment efficiency (defined as the ratio of the enriched ribosome footprint density to the total ribosome footprint density). Nascent chains that interact well with TF show positive log values, whereas those that interact poorly with the chaperone show negative log values. **(B)** A histogram comprising the overall enrichment efficiency of each nascent chain for those with known GO (gene ontology) annotations based on cellular localization (i.e., cytoplasm, GO = 0005737; inner membrane, GO = 0019866; outer membrane, GO = 0009279). The number of genes was represented as a fraction of the total, with the shaded area reflecting the total number. **(C)** Growth analyses of cells expressing or lacking TF. 1:10 serial dilutions (horizontal dimension) of indicated strains (vertical dimension) were spotted on LB plates containing 10 μM IPTG, 50 μg/ml of ampicillin, and specified levels of SDS/EDTA. **(D)** Same as Figure 6C, but dilutions were spotted on LB plates containing 10 μM IPTG, 50 μg/ml of ampicillin, and specified levels of vancomycin. **(E)** Chemical sensitivities of BW25113 Δ*tig* cells correlated with those of more than 3,900 bacterial mutants (Nichols et al., 2011) and represented as a histogram of correlation [R] values. Note that Δ*bamD* was not included in this set. **(F)** Same as Figure 6E, but for BW25113 Δ*yfgC* cells.

Figure 7

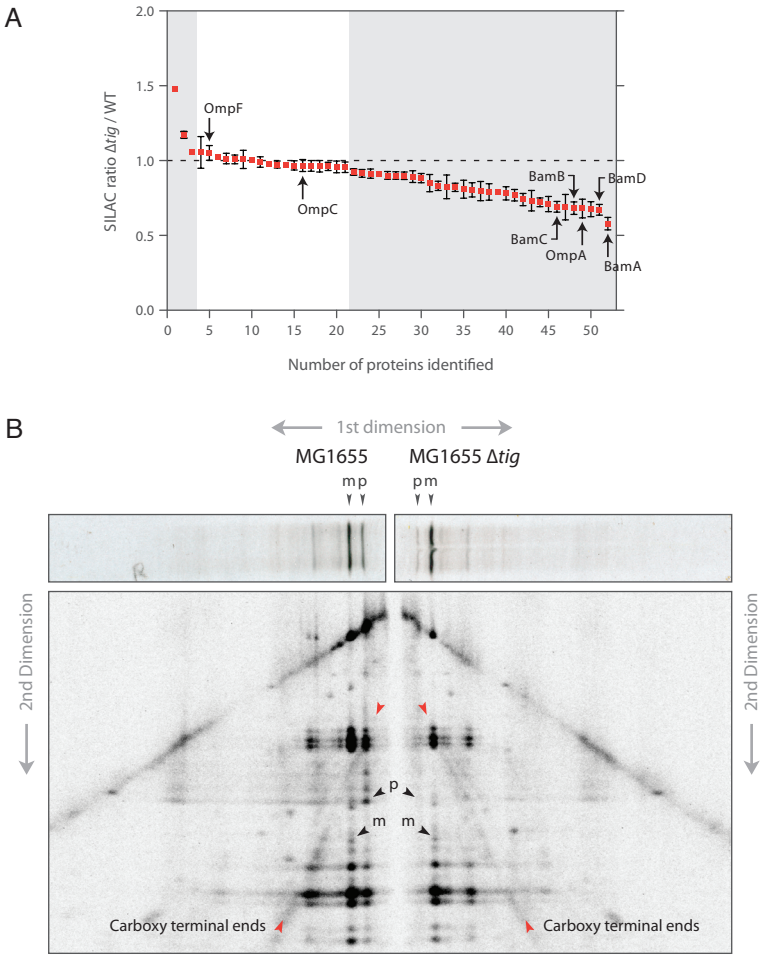

A
SILAC ratio Δtig / WT
Number of proteins identified
OmpF
OmpC
BamB
BamD
BamC
OmpA
BamA
B
1st dimension
MG1655
MG1655 Δtig
mp
pm
2nd Dimension
2nd Dimension
p
m
m
Carboxy terminal ends
Carboxy terminal ends

Figure 7

TF Absence Causes a Broad Reduction in Outer Membrane Protein Levels and Shifts the Mode of Translocation

(A) Quantification of proteins from isolated outer membranes using SILAC. The SILAC ratio (Δ*tig*/wild-type) and corresponding standard error of the mean were calculated for all outer membrane proteins identified with at least three peptides. **(B)** 2D gel assay for monitoring translocation of newly synthesized LamB. Wild-type and Δ*tig::kan* cells were pulse-labeled with ^{35}S-methionine for 30 seconds and quenched using 5% TCA. LamB chains were immunoprecipitated and resolved by 12% SDS-PAGE (first dimension). Gel slices were excised, digested in gel with V8 protease, and resolved by 15% SDS-PAGE (second dimension). Red arrows highlight C-terminal fragments that converge either to the precursor (p) as seen for wild-type cells or mature (m) form as seen for Δ*tig* cells (black arrows).

EXTENDED EXPERIMENTAL PROCEDURES

General Method for Ribosome Profiling in Bacterial Cells

Bacterial ribosome profiling for *Escherichia coli* has been successfully implemented for various laboratory strains including BW25113, MC4100, and MG1655.

Cell Growth and Harvest by Centrifugation

A saturated culture was diluted to an OD_{600} of 0.1 in 1 l of LB media. Cells were grown at 37°C with vigorous shaking until reaching an OD_{600} between 0.4 and 0.5. Intact cells were pre-treated with chloramphenicol to a final concentration of 100 μg/ml for 2 min at standard growth conditions. The culture was poured over ice cubes supplemented with 100 μg/ml of chloramphenicol. Cells were spun down at 4500x*g* for 10 min at 4°C. The cell pellet was washed with 10 ml of pre-chilled resuspension buffer (10 mM $MgCl_2$, 100 mM NH_4Cl, 20 mM Tris pH 8.0, 1 mM chloramphenicol) and spun down at 3000x*g* for 5 min at 4°C. The cell pellet was next resuspended in 7.5 ml of pre-chilled lysis buffer (10 mM $MgCl_2$, 100 mM NH_4Cl, 20 mM Tris pH 8.0, 0.1% NP-40 *or* sodium deoxycholate, 0.4% Triton X-100, 100 U/ml of RNase-free DNase I (Roche), 0.5 U/μl of Superase•In (Ambion), 1 mM chloramphenicol). Resuspended cells were incubated on ice for 5 min and dripped over liquid nitrogen. Frozen cell pellets were cryogenically pulverized (Retsch MM301, 50 ml grinding jar, 25 mm grinding ball) at 15 Hz for 3 min in five sets with canisters re-chilled in liquid nitrogen between each cycle. The pulverized cells were transferred to a 50 ml Falcon and stored at –80°C.

Cell Growth and Harvest by Filtration

A saturated culture was diluted to an OD_{600} of 0.1 in 200 ml of LB media. Cells were grown at 37°C with vigorous shaking until reaching an OD_{600} between 0.4 and 0.5. Cells were filtered using a 0.22 μm nitrocellulose membrane (GE, catalog no.

E02WP09025) in a 90 mm glass filtration system (Kontes), scraped using a pre-warmed scoopula, and immediately submerged in liquid nitrogen. The frozen cells were dislodged using a pre-chilled spatula. 0.65 ml of lysis buffer dripped over liquid nitrogen (10 mM $MgCl_2$, 100 mM NH_4Cl, 20 mM Tris pH 8.0, 0.1% NP-40 *or* sodium deoxycholate, 0.4% Triton X-100, 100 U/ml of RNase-free DNase I (Roche), 0.5 U/μl of Superase•In (Ambion), 1 mM chloramphenicol) was combined with the frozen cells for pulverization (Retsch MM301, 10 ml grinding jar, 12 mm grinding ball) at 15 Hz for 3 min in five sets with canisters re-chilled in liquid nitrogen between each cycle.

Lysate Preparation

Pulverized cells were thawed in a 30°C water bath for 1 to 3 min (or until the lysate completely thawed) and incubated for 10 min in an ice-water bath. The lysate was spun down at 20,000x*g* for 10 min at 4°C. The clarified supernatant was collected and used immediately or frozen in liquid nitrogen and stored at –80°C. The concentration of the clarified lysate was determined by Nanodrop (Thermo Scientific) using a 1:100 dilution in 10 mM Tris pH 7.0.

Ribosome Footprinting

25 Abs_{260} units of the clarified lysate (supplemented with 5 mM $CaCl_2$) were digested with 1,500 enzyme units of micrococcal nuclease (Roche, catalog no. 10107921001) for 1 hr at 25°C and shaken at 1400 rpm in a thermomixer (Eppendorf). Digestions were quenched with EGTA to a final concentration of 6 mM. 1 enzyme unit for MNase is defined as an increase of 0.005 Abs_{260} units per min measured in a SpectraMax plate reader using 10 μg/ml of salmon sperm DNA as substrate supplemented with 5 mM $CaCl_2$ and 10 mM Tris pH 8.0 in a reaction volume of 0.1 ml.

The 30S subunit of *E. coli* ribosomes inhibits RNase I activity by an unknown mechanism (Datta and Burma, 1972). For this reason, we employed MNase, which also

has the advantage of being able to be inhibited by addition of EGTA. We also observed that even for organisms where it is possible to use RNase I, ribosomal RNAs were typically less susceptible to digestion by MNase than by RNase I. This minimized contamination by ribosomal RNA.

10 to 55% (w/v) sucrose gradients (buffered in 10 mM $MgCl_2$, 100 mM NH_4Cl, 20 mM Tris pH 8.0, 1 mM chloramphenicol, 2 mM DTT) were prepared using a BioComp Gradient Master (BioComp Instruments) in polyclear centrifuge tubes (SETON, catalog no. 7030). Both digested and control samples were loaded on gradients and spun in an ultracentrifuge using an SW41 rotor at 35,000 rpm for 2.5 hr at 4°C. Gradients were fractionated using a BioComp Gradient Fractionator with the flow rate set to 0.2 mm per sec. Monosomes were manually collected by following the absorption trace at 254 nm using a BIO-RAD Econo UV monitor.

Library Generation

Ribosome footprints were converted to a cDNA library as previously described (Ingolia, 2010). Monosome fractions were denatured by adding SDS to a final concentration of 1% (w/v). 0.7 ml of denatured monosomes was extracted once with 0.7 ml of hot acid phenol, once with 0.7 ml of acid phenol, and once with 0.6 ml of chloroform; precipitated; and resuspended in 20 μl of 10 mM Tris pH 7.0. 25 μg of RNA (quantified by Nanodrop) was mixed with 2x TBE-urea sample loading buffer (Invitrogen) and resolved on a 15% TBE-urea gel (Invitrogen) in 1x TBE (Ambion) at 200V for 65 min. A band between 28 to 42 bp was excised using 10 bp ladder (Invitrogen) as the standard, gel purified, precipitated, and resuspended in 15 μl of 10 mM Tris pH 7.0. The 3' ends of RNA were dephosphorylated by T4 PNK (NEB) at 37°C for 1 hr; the enzyme was subsequently heat inactivated at 75°C for 10 min after which the RNA was precipitated and resuspended in 10 μl of 10 mM Tris pH 7.0. 5 pmol of RNA quantified by

BioAnalyzer (small RNA kit, Agilent) was diluted to 5 μl in 10 mM Tris pH 7.0 and ligated at 37°C for 2.5 hr to 1 μl of 1 μg/μl Linker-1 (5'_App/CTGTAGGCACCATCAAT/3ddC_3', IDTDNA) supplemented with 8 μl of 50% sterile filtered PEG MW 8000, 2 μl of 10x T4 RNA ligase buffer (NEB), 2 μl of 100% DMSO (Sigma), 1 μl of 20 U/μl Superase•In (Ambion), and 1 μl of T4 ligase 2, truncated (NEB). The ligated products were precipitated and resolved on a 10% TBE-urea gel (Invitrogen) in 1x TBE at 200V for 50 min. A band between 45 and 60 bp was excised using 10 bp ladder as the standard, gel purified, precipitated, and resuspended in 10 μl of 10 mM Tris pH 7.0. The gel extracted products were reverse transcribed with no more than five molar excess of Linker-1 (5'_5phos/GATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CACTCA/iSp18/CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG_3', IDTDNA) using Superscript III (Invitrogen) in a 20 μl reaction volume at 50°C for 30 min. RNA products were hydrolyzed by the addition of NaOH (at a final concentration of 0.1 mM) and incubated at 95°C for 15 min. The reverse transcribed cDNA was resolved on a 10% TBE-urea gel in 1x TBE for 70 min at 200V. A band between 125 to 150 bp was excised using 10 bp ladder as the standard, gel purified, and resuspended in 15 μl of 10 mM Tris pH 8.0. The cDNA was circularized with CircLigase (EPICENTRE) in a 20 μl reaction volume at 60°C for 1 hr; the enzyme was heat inactivated at 80°C for 10 min after which the circDNA was stored at –30°C for further use.

5 μl of circDNA was PCR amplified with 5'_AATGATACGGCGACCACCGA_3' and 5'_CAAGCAGAAGACGGCATACGA_3' using Phusion polymerase (NEB) for 7 to 10 cycles. Amplicons were resolved on an 8% polyacrylamide gel (Invitrogen) in 1x TBE at 180V for 45 min. A band between 125 to 150 bp was excised using 10 bp ladder as the standard, gel purified, precipitated, and resuspended in 10 μl of 10 mM Tris pH 8.0. The PCR amplified DNA library was quantified by BioAnalyzer (high sensitivity DNA kit,

Agilent) and sequenced by Illumina Genome Analyzer II. 10 to 13 pM solution was used for generating clusters with 5'_CGACAGGTTCAGAGTTCTACAGTCCGACGAT_3' as the sequencing primer.

Selective Ribosome Profiling in Bacterial Cells

Bacterial Strains and Plasmids

MC4100 Δ*tig::kan* and MG1655 Δ*tig::kan* were constructed by P1 transduction with strain specific P1-lysates. All *tig* encoding plasmids used in this study were derived from pTrc99B (Amann et al., 1988), enabling *tig* expression under the control of an IPTG-inducible hybrid trp/lac promoter. pTrc99 was used to construct pTrc-Tig and pTrc-TigAAA. pNde, which is a derivative of pTrc99 that contains an NdeI restriction site, was used to construct pTig, pTig-Avi, pTig-TEV-Avi, and pTigAAA-TEV-Avi. pNde derivatives required higher concentrations of IPTG to achieve endogenously expressed levels of TF when compared with pTrc99 based plasmids. Cloning of plasmids expressing C-terminally tagged TF was achieved by PCR, whereby the 3' end of the *tig* open reading frame was fused in frame with an AviTag (GSGLNDIFEAQKIEWHE) or a version cleavable by TEV protease (GSGENLYFQSGRSGLNDIFEAQKIEWHE).

MG1655 Δ*tig::tig-TEV-AviTag* Δ*ompF::cm* was constructed by P1 transduction using strain specific P1-lysates. *ompF* derived plasmids were cloned between the EcoRI and KpnI restriction sites of pRC10 (Chaba et al., 2007), itself a derivative of pTrc99A (Amann et al., 1988). *ompF ss** had mutations at the following residues: N5D, V9E, A13D, and V16E. *ompF Δ2-48* lacked codons 2 through 48, whereas *ompF Δ2-96* lacked codons 2 through 96. *ompF myo[1-50]* had 50 codons of human myoglobin inserted immediately following the signal sequence, but codon R54 was mutated by site directed mutagenesis from R(AGG) to R(CGG).

IPTG Titration Experiments

Cultures of MC4100 Δ*tig::kan* transformed with pTig-TEV-Avi were grown in LB media supplemented with 100 μg/ml of ampicillin and indicated concentrations of IPTG. MC4100 grown in LB was used as the wild-type control. Cells were treated with chloramphenicol to a final concentration of 100 μg/ml when they reached an OD_{600} of 0.45 and chilled on ice. Cells were next harvested by centrifugation, resuspended in SDS sample buffer, and boiled for 5 min. Equal amounts were subjected to SDS-PAGE and Western blotting using polyclonal antiserum specific for TF (lab collection).

Real-Time PCR

Cultures of MG1655 Δ*tig::tig-TEV-AviTag* Δ*ompF::cm* with plasmids containing *ompF* derivatives were grown in LB media supplemented with 100 μg/ml of ampicillin and 200 μM IPTG. 4 ml cultures were harvested by centrifugation and flash frozen. 0.3 ml of lysis solution (0.5% SDS, 30 mM NaAOc pH 5.5, 10 mM EDTA) was next added to the frozen cell pellet. Total RNA was extracted twice with acid phenol, once with chloroform, ethanol precipitated, and resuspended in 10 mM Tris 7.0. 10 μg of total RNA was digested with 2 μl of DNase I (NEB) in 0.1 ml for 10 min at 37°C. 1 μl of 0.5 M EDTA was mixed in each digested sample and incubated for 10 min at 37°C. Total RNA was further purified using the RNA Cleanup Protocol from the QIAGEN RNeasy Mini Kit. 1 μg of purified total RNA was combined with 2 μl of 50 μM random hexamer solution (ABI), 1 μl of 20 U/μl Superase•In (Ambion), 1 μl of 10 mM dNTPs (NEB), 2 μl of 10x M-MuLV buffer (NEB), and 1 μl of 200 U/μl M-MuLV reverse transcriptase (NEB). This mixture was incubated at room temperature for 10 min, next at 42°C for 1 hr, and finally at 90°C for 10 min. Real time PCR was performed using the DyNAmo HS SYBR Green qPCR Kit (Finnzymes) on a DNA Opticon Real Time Cycler (MJ Research). Primers were designed using SciTools (IDTDNA) for *ompF* and *ssrA*.

Cell Growth and Harvest for Sequencing and Lysis

Overnight cultures of MC4100 Δ*tig::kan* (transformed with indicated plasmids) were diluted to an OD_{600} of 0.02 in fresh LB media supplemented with 100 μg/ml of carbenicillin, 40 μg/ml of biotin, and 10 μM IPTG for pTrc-Tig *or* 70 μM IPTG for pTig-Avi, pTig-TEV-Avi, and pTig-AAA-TEV-Avi. Diluted cultures were grown at 37°C to an OD_{600} of 0.45 with vigorous shaking.

To harvest cells by centrifugation, 500 ml cultures were initially incubated with chloramphenicol to a final concentration of 100 μg/ ml and shaken for 2 min at 37°C. To quickly chill the cells, cultures were poured over ice cubes of equivalent volume containing 100 μg/ml of chloramphenicol. Cells were then harvested by centrifugation at 4500x*g* for 10 min at 4°C, followed by one wash in ice cold buffer (50 mM HEPES pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM chloramphenicol). The pellet was resuspended in 3 ml of ice cold crosslinking/lysis buffer (50 mM HEPES pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM chloramphenicol, 1 mM PMSF, 0.4% Triton X-100, 0.1% NP-40, 100 U/ml of RNase-free DNase I) and flash frozen by dripping the cell suspension into liquid nitrogen. Frozen cell pellets were pulverized for cell lysis in a 50 ml stainless steel grinding jar (MM301/400, 25 mm grinding ball; Retsch) for 5 cycles, each at 15 Hz for 3 min.

To harvest cells by filtration, 200 ml cultures were filtered using a pre-warmed 90 mm glass filtration system (Kontes) with nitrocellulose membranes of 0.2 μm in pore size (MicroSep Cellulosic, GE). Harvesting by filtration was half as efficient in collecting cells as harvesting by centrifugation, thus 1 l was cultured and filtered five independent times with each filter used once for 200 ml. Following the filtration of growth media, cells were quickly scraped off the membrane with a pre-warmed scoopula and immediately submerged in liquid nitrogen for flash freezing. Frozen cell pellets (from 200 ml of

culture) were combined with 650 ml of frozen crosslinking/lysis buffer and subsequently pulverized in a 10 ml mixer mill jar using aforementioned milling conditions.

Crosslinking TF to Ribosome-Nascent Chain Complexes

Frozen cell powder was added in batches to the crosslinking/lysis buffer containing DSP (dithiobis succinimidyl propionate; Pierce) with constant stirring at room temperature, which roughly took 10 min in span. The amount of crosslinking/lysis buffer was equivalent to the volume of resuspended cells before pulverization, with the final concentration of DSP adjusted to 2.5 mM. Once the cell powder was completely dissolved, the thawed lysate was further incubated for 5 min. The crosslinking reaction was quenched with Tris pH 8.3 to a final concentration of 100 mM and further incubated for 5 min at room temperature with constant stirring. The quenched reaction was chilled on ice for 10 min before the lysate was clarified by centrifugation at 20,000x*g* for 10 min at 4°C.

Chemical crosslinking by EDC (1-ethyl-3-[3-dimethylaminopropyl] carbodiimide; Pierce) was done as described above for DSP crosslinking, except EDC was used at a final concentration of 20 mM pH 5.8 and the reaction was quenched with 250 mM glycine, 100 mM Tris 8.0, and 4 mM $NaHCO_3$.

MNase Digestion of Crosslinked Lysates

Clarified lysates were quantified by Nanodrop (Thermo Scientific). 25 Abs_{260} units (corresponding to 1 mg of total RNA) were taken for the undigested control. 450 Abs_{260} units (corresponding to 18 mg of total RNA) were subjected to nuclease digest for 1 hr at 25°C using 20250 enzyme units of MNase supplemented with 40 μl of Superase•In (Ambion). Reactions were quenched with EGTA pH 8.0 to a final concentration of 6 mM and chilled on ice.

Isolating Monosomes by Sucrose Density Gradients

10 to 55% (w/v) sucrose density gradients were prepared as described for general ribosome profiling. Monosomes from nuclease treated samples were collected for the total control by sucrose density gradients, except DTT was omitted from the gradient buffer.

Affinity Purification of TF Crosslinked RNCs

Nuclease treated lysates were loaded on 25% sucrose cushions (50 mM Tris pH 7.5, 1 M NaCl, 10 mM $MgCl_2$, 1 mM chloramphenicol, 1 mM PMSF) and spun at 45,000 rpm for 4.5 hr at 4°C in a Ti45 rotor (Beckman). As soon as the supernatant was discarded, the pellet was quickly washed with wash buffer (50 mM Tris pH 7.0, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM chloramphenicol, 1 mM PMSF, 0.4% Triton X-100, 0.1% NP-40). The ribosomal pellet was resuspended in 1.25 ml of wash buffer and incubated overnight on ice.

For each sample, 220 μl of the matrix (50% slurry of Strep-Tactin sepharose; IBA, Göttingen, Germany) was equilibrated in wash buffer. The resuspended ribosome fraction was incubated with the matrix for 1 hr on an overhead roller at 4°C. The matrix was centrifuged at 500x*g* for 1 min. The supernatant was carefully aspirated by pipette, removing unbound ribosomes. The matrix was washed three times in wash buffer for 45 min at 4°C on an overhead roller, followed by centrifugation at 500x*g* and aspiration of the supernatant by pipette. The matrix was next washed once with cleavage buffer (50 mM Tris pH 7.0, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM chloramphenicol).

130 μl of cleavage buffer plus 5.8 μM TEV protease (lab collection) was added to the matrix for specific elution of crosslinked TF-RNCs and incubated for 30 min at room temperature on an overhead roller. The matrix was centrifuged at 500x*g* after which the supernatant was collected and saved. TEV cleavage was repeated using the conditions described above. 120 μl of cleavage buffer was added to the matrix to capture residual

TF-RNCs trapped in the slurry. The collected eluates were pooled for subsequent isolation of mRNA footprints (see *Library generation* under General ribosome profiling section). Typical yields per sample ranged between 50 to 100 μg of RNA as measured by Nanodrop. The affinity matrix was washed once with wash buffer for 30 min, and another time with cleavage buffer and saved for gel analyses.

Crosslinking Analysis by Polyacrylamide Gel Electrophoresis and Immunoblotting

Samples were mixed with SDS sample buffer (625 mM Tris pH 6.8, 50% glycerol, 15% β-mercaptoethanol, 0.05% bromophenol blue) unless indicated differently and resolved on a 12% SDS polyacrylamide gel. For nonreducing SDS-PAGE, sample buffer lacking reducing agents was used. SDS gels were visualized by standard methods for silver staining.

For immunoblotting, samples were first subjected to SDS-PAGE, then electro-transferred to PVDF membranes (Roth, Germany) and probed with indicated antisera (lab collection). Alkaline phosphatase conjugated to secondary antibodies were used for ECF detection using an FLA-3000 PhosphoImager (Fuji Photo Film, Tokyo, Japan). Bands were quantified using Multi-Gauge (Fuji Photo Film, Tokyo, Japan).

Sequencing Analysis

Sequences were imaged by Illumina Genome Analyzer II and analyzed by GAPipeline v.0.3.0; sequences were next aligned to the bacterial genome using Bowtie v.0.12.0 as previously described (Ingolia, 2010; Ingolia et al., 2009). NC_012759.fna (RefSeq) was used for MC4100. Matrix and phasing parameters were determined from ϕX DNA. Bacterial genomes were obtained at NCBI Reference Sequence Bank (http://www.ncbi.nlm.nih.gov/RefSeq).

Assigning Footprints to the Genome

Genome-aligned reads with more than two mismatches were excluded. Moreover, only uniquely aligned sequencing reads were included to facilitate the analysis; thus it was assumed that nonunique alignments (i.e., sequences that matched to multiple places in the genome exactly at the nucleotide level) came from noncoding sequences, such as rRNA and tRNA molecules.

Because footprint reads varied in length, scores were fractioned and distributed across internal residues, in contrast to footprints derived from eukaryotic systems in which the 50 position was scored once for each sequencing read (Guo et al., 2010; Ingolia et al., 2009). For footprints between 25 and 40 nt in length, 12 nt were trimmed from each end with the remaining residues given a score of 1/N in which N equals the number of positions leftover after discarding the 5' and 3' ends. By blurring the signal across the central residues, the certainty to which the ribosome could be positioned was lowered by one to five codons (Figures 1C through 1E).

Quantifying Ribosome Density

To measure the overall expression rate for each gene (expressed as the density across an ORF in reads per kilobase million), the number of reads aligned to an ORF was normalized by its overall length and the total number of reads aligned to all ORFs. When comparing expression rates for each gene between replicates (Figure 1F), the sum of reads between them had to exceed 100 counts, ensuring that measurements were reliable and not dictated by sampling error (Ingolia et al., 2009). When comparing expression rates between different genes of the same operon (Figure 1G), each gene required at least 100 counts. The list of transcription units were found at http://ecocyc.org (Keseler et al., 2009) for MG1655 and applied to MC4100.

To measure ribosome density across each gene, the reads at each position were normalized by the total number of reads aligned to all ORFs only (i.e., in units of reads

per kilobase). Ribosome densities were visualized by MochiView v.1.45 (Homann and Johnson, 2010), an open source genome browser. The genes identified in Figures 1D and 1E were found by browsing MochiView, indicating that other unidentified short ORFs are likely to exist.

Meta-Gene Analyses

The average ribosome density across nucleotides 280 to 400 was used as the mean ribosome density for each gene. To obtain the meta-gene profile illustrated in Figure 1B, the ribosome density profile for each well-expressed gene (i.e., those with an overall density greater than 0.08 reads per base) was scaled by its own mean ribosome density. This gave differently expressed genes similar expression rates (i.e., equal weighting). Each mean normalized ribosome density profile was next aligned by its start codon (or by its stop codon) and averaged across each position—that is if the transcript was long enough to be included for averaging. Thus the output was an average of numerous mean normalized ribosome density profiles.

Measuring Enrichment Efficiency

To generate meta-gene profiles illustrated in Figures 3A and 4B, meta-gene ribosome density profiles were separately computed for both footprints derived from affinity purified TF-RNCs and those derived from all footprints. But ORFs encoding less than 135 codons were excluded from this analysis, because their inclusion would inflate the overall range in enrichment efficiency near the N-terminus. In addition, 100 codons from the 5' end were trimmed and excluded when deriving meta-gene ribosome density profiles starting from the stop codon only. Ratios between the two meta-gene read density profiles were taken across each position (i.e., enriched over total), yielding the average mean normalized enrichment efficiency.

Individual enrichment efficiency profiles shown in Figure 3B were determined for

each gene by taking the ratio across the coding sequence between the ribosome density profile computed for affinity purified footprints and the ribosome density profile computed for the total collection of footprints. However, these density profiles were scored differently than previously described, with the score at each position also containing the scores of ± 20 nucleotides from it. This was done because many genes were marked by regions lacking continuous density.

To assess initial chaperone engagement for each gene as in Figures 3C and 4C, individual enrichment efficiency profiles were computed for very well-expressed genes, here defined as those having a read density of more than 1.5 reads per base. After measuring when each enrichment efficiency profile first crossed a threshold of 1.3 arbitrary units, the collection of these lengths (in codons) was represented as a histogram. Genes that did not meet this threshold were excluded from this analysis. Genes were also excluded if they already exceeded the threshold at the first codon, which accounted for less than 1% of ORFs.

Gene-by-gene correlations across enrichment efficiency profiles (from very well-expressed genes with a read density of more than 1.0 read per base) were determined between DSP replicates or DSP and EDC experiments as in Figure 4D.

To measure the overall enrichment efficiency as in Figure 6A, the sum of the enriched footprint density was divided by the sum of the total footprint density per gene, but excluding 100 codons from the 5' end. Genes were separated based on their GO (gene ontology) annotation (i.e., GO = 0005737, GO = 0019866, GO = 0009279) and compared with the total set using the rank sum test from MatLab v.7.5.0. GO annotations were found at http://ecocyc.org (Keseler et al., 2009) for MG1655 and converted for MC4100.

In Vitro and In Vivo Analyses

In Vitro Analysis of N-Terminal Enzymatic Processing of Nascent Chains

S135 translation extracts were prepared from MC4100 Δ*tig* cells using a modified version of an established protocol (Müller and Blobel, 1984). Cultures were grown at 37°C in S135 medium (0.9% tryptone/peptone, 0.08% yeast extract, 0.56% NaCl, 1 ml/l of 1 M NaOH, 0.08% glucose) to an OD_{600} of 1.2 and harvested by centrifugation. Cell pellets were resuspended in S135 buffer (10 mM triethanolamine (TEA) pH 7.5, 14 mM $Mg(OAc)_2$, 60 mM KOAc) containing 1 mM dithiothreitol (DTT) and 0.5 mM phenyl methane sulfonyl fluoride (PMSF) [1 ml of S135 buffer per 1 g of wet pellet]. Cells were lysed using a French Pressure Cell. Lysates were next cleared by centrifugation at 30,000x*g* for 30 min at 4°C. 60 µl of 1 M TEA pH 7.5, 0.6 µl of 1 M DTT, 1.6 µl of 1 M $Mg(OAc)_2$, 6 µl of 1 mM each of all amino acids except methionine, 6 µl of 1 mM methionine, 2 µl of 0.25 M ATP, 27 µl of 0.2 M phosphoenolpyruvate, and 2.4 µl of 2 mg/ml pyruvate kinase were mixed per ml of supernatant (also known as S30 extract) and incubated for 1 to 2 hr at 37°C. Extracts were cooled on ice and dialyzed three times using a 12 to 14 kDa cut-off against S135 buffer containing 1 mM DTT. Following dialysis, 1 ml aliquots were ultracentrifuged in a TLA-100.2 rotor at 135,00x*g* for 13 min at 4°C. 750 µl of supernatant from each aliquot were combined and dialysed three times against S135 buffer lacking DTT but containing 1 mM tris(2-carboxyethyl)phosphine (TCEP). The final S135 extract was aliquoted, snap-frozen in liquid nitrogen, and stored at –80°C.

Each translation reaction (in a total volume of 25 µl) contained 2.5 µl of compensation buffer (392 mM TEA pH 7.5, 1.35 M KOAc, 99 mM $Mg(OAc)_2$, 8 mM spermidine); 6.25 µl of power mix (12.8% polyethylene glycol MW 6000, 48 mM phosphoenolpyruvate, 32 mM creatine phosphate, 1 mM TCEP, 0.16 mM each of all

amino acids except methionine, 10 mM ATP, 2 mM each of GTP, CTP, and UTP; buffered at pH 7.0); 40 μg/ml of creatine phosphate kinase; 300 ng of purified template DNA; 0.5 μl of purified T7 polymerase; 2 μl of S135 extract; 0.5 μl of ^{35}S-methionine at 10 μCi/l; 100 nM $CoCl_2$; and specified concentrations of PDF and MAP. Reactions were incubated in a thermomixer shaken at 900 rpm for 30 min at 37°C. Control reactions were supplemented with actinonin to a final concentration of 30 ng/ml. All reactions were terminated by adding 35 μl of ice cold 10% TCA. Samples were incubated on ice for 30 min and centrifuged at 20,000x*g* for 30 min at 4°C. Precipitants were resuspended in alkaline sample buffer, run on 11% Tris-tricine gels, stained by Coomassie Brilliant Blue, and dried before the autoradiogram was developed.

In Vivo Analysis of TF Overexpression on the Processing of Nascent Polypeptide Chains

MC4100 Δ*acrA::kan* cells were transformed with pTrc99, pTrc-Tig, or pTrc-TigAAA. Serial dilutions of overnight cultures were spotted on LB plates containing 100 μg/ml of ampicillin and indicated concentrations of IPTG and/or actinonin, and incubated for 16 hr at 37°C.

In Vivo Complementation Analysis

MC4100 Δ*tig::kan* Δ*dnaK52* cells (Vorderwülbecke et al., 2004) were transformed with pNde, pTig, or pTig-TEV-Avi. Cells were grown on selective LB agar plates containing 70 μM IPTG for 21 hr at 30°C to facilitate the episomal expression of *tig* or *tig-TEV-Avi*.

Colonies were struck to singles twice on LB plates containing 100 μg/ml of ampicillin and 70 μM IPTG before they were inoculated and grown overnight at 30°C in LB media containing 100 μg/ml of ampicillin and 70 μM IPTG. 1:10 serial dilutions were spotted on LB plates containing 100 μg/ml of ampicillin and 70 μM IPTG, and incubated

for 17 hr at 34°C, 21 hr at 30°C, or 43 hr at 25°C.

SDS/EDTA and Vancomycin Sensitivity Assay

As shown in Figures 6C and 6D, MG1655 cells harboring pTrc99 (empty vector) and the corresponding MG1655 Δ*tig::kan* mutant cells containing pTrc99 or pTrc-Tig were grown to stationary phase in LB media containing 100 μg/ml of ampicillin. 1:10 serial dilutions were spotted on LB plates containing 50 μg/ml of ampicillin, 10 μM IPTG, and indicated concentrations of SDS/EDTA or vancomycin. Plates were incubated for 16 hr at 37°C.

As shown in Figure S5A, MG1655 cells harboring pNde (empty vector) and the corresponding MG1655 Δ*tig::kan* mutant cells containing pNde, pTig, or pTig-TEV-Avi were grown to stationary phase in LB media containing 100 μg/ml of ampicillin. 1:10 serial dilutions were spotted on LB plates containing 50 μg/ml of ampicillin, 70 μM IPTG, and indicated concentrations of SDS/EDTA or vancomycin. Plates were incubated for 16 hr at 37°C.

To measure growth curves (shown in Figures S4 and S5B), stationary cultures were diluted in fresh LB media and grown to an OD_{600} of 0.4. Cells containing plasmids were supplemented with IPTG. Cultures were diluted to an OD_{600} of 0.1 in LB media containing indicated concentrations of SDS/EDTA or vancomycin. Optical density was measured at 600 nm following dilution at indicated time points.

Analysis of Chemical Genetics Screen

Correlation values across various strains were taken from a previously published chemical genetic screen (Nichols et al., 2011) and represented as a histogram.

Analysis of σ^E Activity

σ^E activity was determined by measuring β-galactosidase expression using a

previously described σ^E-dependent reporter system (Ades et al., 1999; Mecsas et al., 1993) in which $\phi\lambda$(*rpoHP::lacX*) was integrated into MG1655 and MG1655 Δ*tig::kan* cells. Cultures were first grown in LB media starting from an OD_{600} of 0.01 at 30°C and β-galactosidase activity was measured at multiple growth phases.

Isolation of Outer Membrane Proteins for SILAC

Light and heavy M9 minimal media were prepared containing 43 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 10 μg/ml of thiamine. Each amino acid was supplemented to a final concentration of 0.5 mM, except Lysine-8:HCl and Arginine-10:HCl (Silantes, Munich, Germany) were used for heavy media. pH was adjusted to 7.0 using NaOH, and glucose was added to a final concentration of 0.4%. MG1655 and MG1655 Δ*tig::kan* cells were grown to stationary phase in both light and heavy M9 minimal media. Stationary cultures were diluted to an OD_{600} of 0.02 in 50 ml of M9 minimal media and grown to mid-log phase at 37°C. Translation was stopped by adding 100 μg/ml of chloramphenicol. Cultures were chilled on ice for 20 min. Equal amounts (gauged by optical density) of wild-type and Δ*tig::kan* cells grown in different media were mixed and harvested by centrifugation. Cell pellets were subsequently frozen in liquid nitrogen.

Frozen cells were thawed in a 30°C waterbath for 2 min, resuspended in lysis buffer (18 mM Tris pH 7.5, 18 mM KCl, 0.36 mM EDTA pH 8.0, 180 μg/ml of lysozyme, 18 μg/ml of RNase A, 9 μg/ml of DNase I), and lysed by sonication. Unlysed cells were cleared by light centrifugation at 500x*g* for 15 min at 4°C. The supernatant (or total lysate) was collected and spun at 13,200 rpm for 16 min. The resulting pellet (or total membrane fraction) was washed once with 20 mM $NaPO_4$ pH 7.0 and three additional times with 20 mM $NaPO_4$ pH 7.0 containing 0.5% sarcosyl, which solubilizes and

removes the inner membrane fraction. The final pellet (or outer membrane fraction) was resuspended in 20 mM $NaPO_4$ pH 7.0.

Mass Spectrometry Analysis

Proteins from the outer membrane fraction were separated by SDS-PAGE. The entire lane was cut in 3 mm segments. Excised gels were reduced by DTT, alkylated by iodoacetamide, and digested by trypsin (Catrein et al., 2005) using a Digest Pro MS liquid handling system (Intavis). Following digestion, the tryptic peptides were extracted from the gel slices with 50% acetonitrile/0.1% trifluoroacetic acid (TFA), concentrated in a SpeedVac vacuum centrifuge until nearly dried, and diluted to a max volume of 30 μl with 0.1% TFA. 25 μl of sample was analyzed by a nanoHPLC system (Tempo nanoMDLC, Applied Biosystems) coupled to an ESI LTQ Orbitrap mass spectrometer (Thermo Fisher). Samples were next loaded on a C18 trapping column (Reprosil-pur, C18, Dr Maisch) with a flow rate of 10 μl/min in 0.1% TFA. Peptides were eluted and separated on an analytical column (75 μm × 150 mm packed with Reprosil-pur, C18, Dr Maisch) with a flow rate of 200 nl/min in a gradient of buffers A (0.1% formic acid) and B (0.1% formic acid and acetonitrile): 0 to 6 min, 3% of B; 6 to 60 min, 3%–40% of B; 60 to 65 min, 60%–90% of B. The column was connected to a nano-ESI emitter (New Objectives). 1500V were applied by liquid junction. One survey scan (resolution at 60000) was followed by five information dependent product ion scans in the LTQ. Only doubly and triply charged ions were selected for fragmentation. Data analysis was performed using MaxQuant (1.2.0.11) containing the search engine Andromeda (Cox et al., 2011). ncbi.Ecoli_K12_substrDh10B.25-Jan-2010, provided by MaxQuant, was used as the database. Trypsin specificity (one missed cleavage), fixed modification of cysteines (carbamidomethyl), and variable modification of methionines (oxidation) were used. The false discovery rate of peptides and proteins was set to 0.01. Proteins with

less than three peptides detected were omitted in this analysis. The outer membrane fraction was normalized to the total fraction, which in turn was normalized by MaxQuant. Outer membrane proteins were identified based on their GO annotation.

Pulse-Labeling and Two-Dimensional Gel Analysis

Cultures of MG1655 and MG1655 Δ*tig::kan* were grown to mid-log phase in M63 minimal media containing 0.2% maltose and an 18 amino acid mix lacking cysteine and methionine. Cultures were incubated with ^{35}S-labeled methionine for 30 seconds. Labeling was terminated by adding cold trichloroacetic acid (TCA) to a final concentration of 5%. Immunoprecipitation and two-dimensional gel analyses were performed as previously described (Randall, 1983).

Figure S1

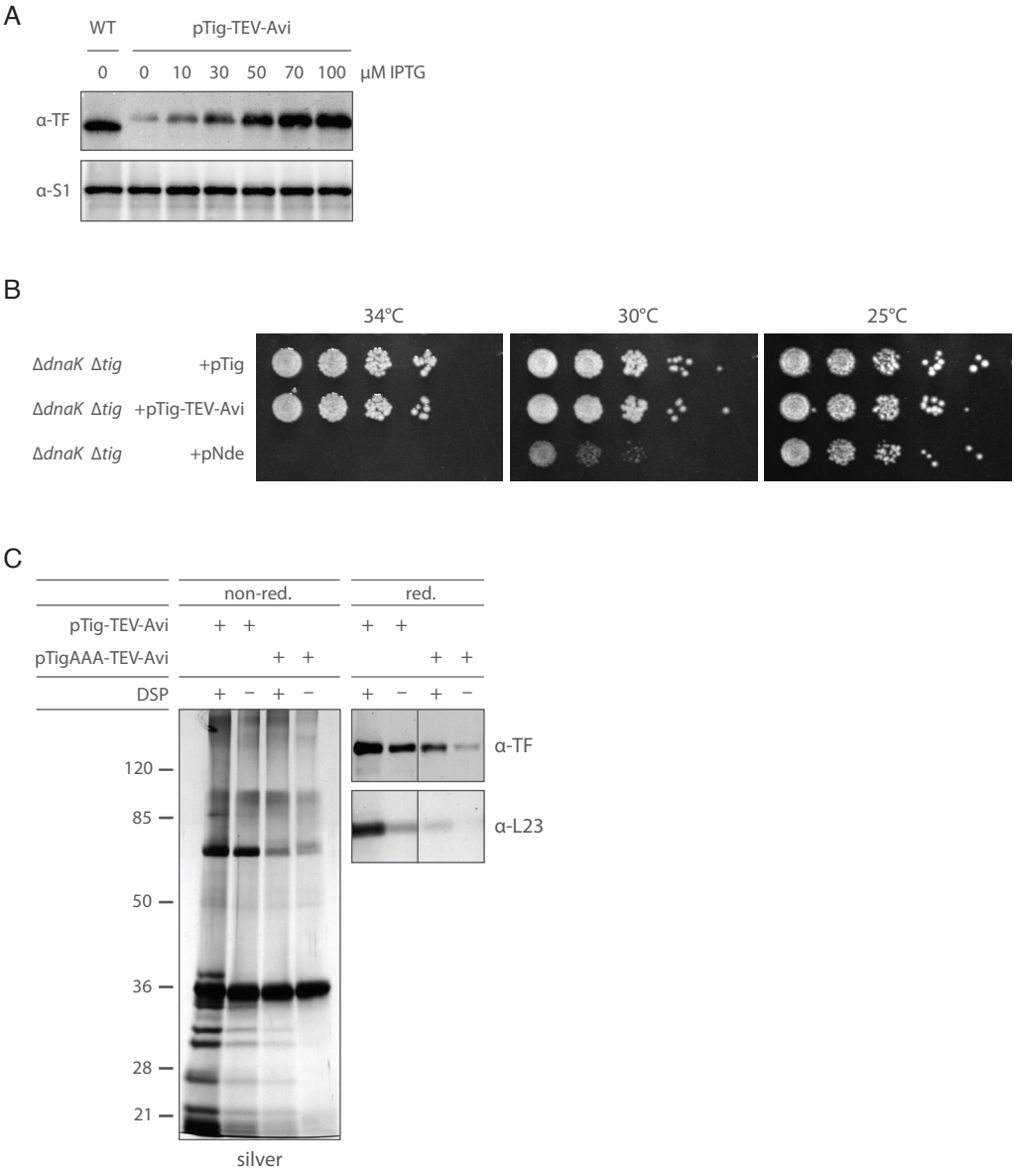
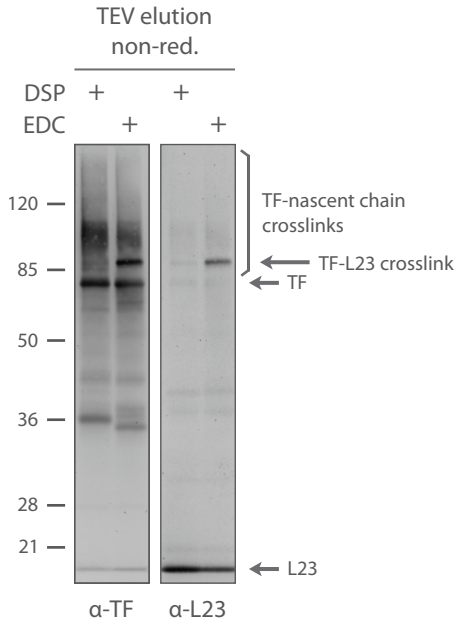

A
WT
pTig-TEV-Avi
0 0 10 30 50 70 100 µM IPTG
α-TF
α-S1
B
34°C
30°C
25°C
ΔdnaK Δtig +pTig
ΔdnaK Δtig +pTig-TEV-Avi
ΔdnaK Δtig +pNde
C
non-red.
red.
pTig-TEV-Avi
pTigAAA-TEV-Avi
DSP
120
85
50
36
28
21
silver
α-TF
α-L23

Figure S1

Characterizing the Affinity Purification Strategy for TF Crosslinked RNCs

(A) IPTG induced expression of plasmid encoded TF matches endogenously expressed levels. Total lysates from wild-type and Δ*tig::kan pTig-TEV-Avi* cells grown to mid-log phase in various IPTG concentrations were resolved by 12% SDS-PAGE under reducing conditions. TF steady-state levels were analyzed by immunoblotting using TF antiserum. The membrane was subsequently stripped and developed using S1 antiserum as the loading control. **(B)** Untagged and TEV-Avi-tagged TF complement the simultaneous knockout of Δ*tig* Δ*dnaK*. Δ*tig* Δ*dnaK* cells transformed with pNde, pTig, or pTig-TEV-Avi were serially diluted and spotted on LB plates containing 100 μg/ml of ampicillin and 70 μM IPTG. Cells were grown for 17 to 43 hr at indicated temperatures. **(C)** Ribosome binding is required for TF-nascent polypeptide interaction. Δ*tig::kan* cells transformed with either pTig-TEV-Avi or pTigAAA-TEV-Avi (encoding a TF mutant deficient in ribosome binding) were grown to mid-log phase and harvested by centrifugation. Lysis, crosslinking, affinity purification, and TEV elution of TF-RNCs were achieved as described in Figure 2. TEV eluates were split in half and analyzed by either non-reducing (non-red.) or reducing (red.) SDS-PAGE. Nonreducing gels were subjected to silver staining, and reducing gels were immunoblotted using TF or L23 antiserum.

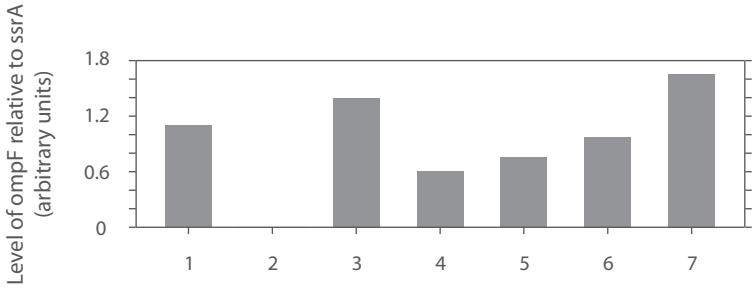
Figure S2



Crosslinking Analysis of Purified TF-RNCs after Filtering Cells

Samples were treated as described in Figure 4A, except cells were not pre-treated with chloramphenicol before harvest, but rapidly filtered and flash frozen.

Figure S3

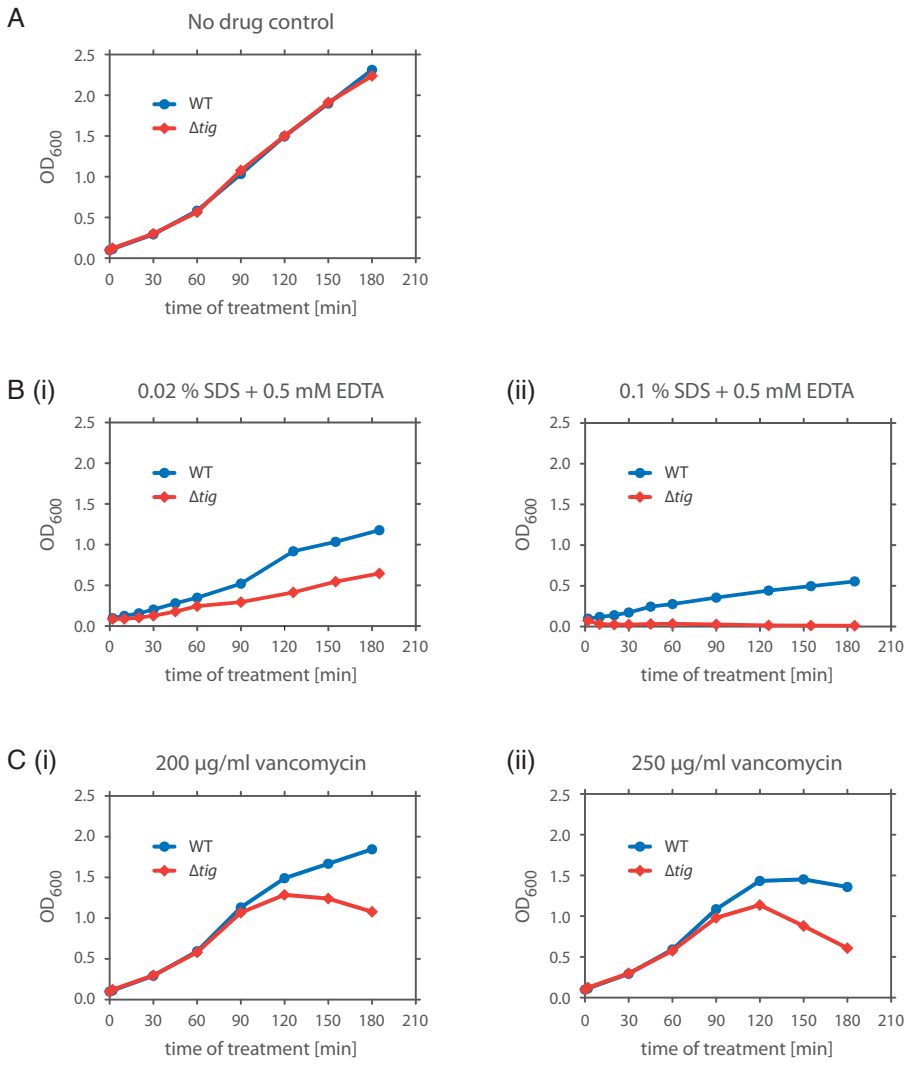


Sample number	Genetic background	Insert	Induced?
1	WT	none	no
2	Δ*ompF*	ompF	no
3	Δ*ompF*	ompF	yes
4	Δ*ompF*	ss*	yes
5	Δ*ompF*	Δ2-48	yes
6	Δ*ompF*	Δ2-96	yes
7	Δ*ompF*	myo insert	yes

Quantifying Expression Levels for ompF Derivatives

MG1655 Δ*tig::tig-TEV-Avi* (WT) or MG1655 Δ*tig::tig-TEV-Avi* Δ*ompF::cm* (Δ*ompF*) cells transformed with pRC10 containing indicated inserts were induced with IPTG unless otherwise noted. Cells were grown to an OD_{600} of 0.5 and harvested by centrifugation. Total RNA was purified and converted to cDNA using M-MuLV reverse transcriptase and random hexamers. Real time PCR was performed using the DyNAmo HS SYBR Green qPCR Kit. ssrA levels were measured for internal normalization.

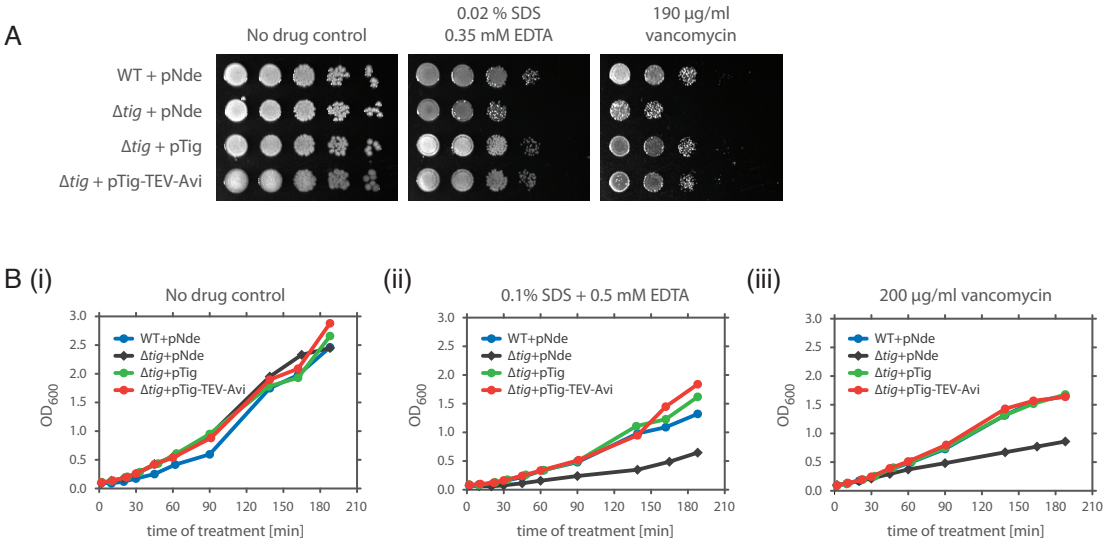
Figure S4



Cells Lacking TF Show Impaired Outer Membrane Integrity

Wild-type and Δ*tig::kan* cells were grown to an OD_{600} of 0.4, and diluted 1:4 in LB media **(A)** only or containing indicated concentrations of **(B)** SDS/EDTA or **(C)** vancomycin. Cell growth was measured by determining OD_{600} at indicated time points.

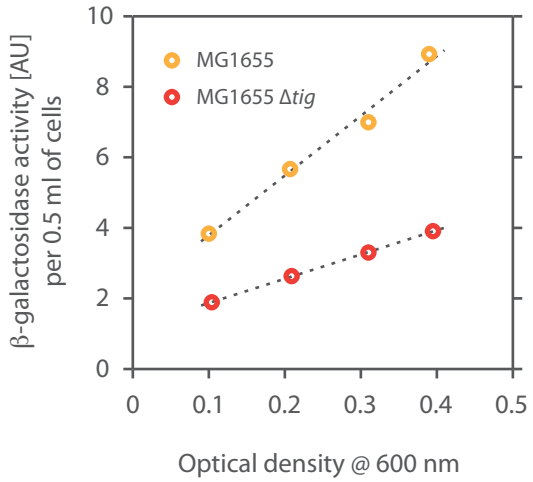
Figure S5



Epitope-Tagged TF Complements the Chemical Sensitivities of Δ*tig* Cells

(A) Stationary cultures of wild-type cells transformed with pNde as well as Δ*tig::kan* cells transformed with pNde, pTig, or pTig-TEV-Avi were spotted as 1:10 serial dilutions on LB plates containing 50 µg/ml of ampicillin, 70 µM IPTG, and specified levels of drugs. **(B)** Strains described in (A) were grown to an OD_{600} of 0.4 in LB media containing 100 µg/ml of ampicillin and 70 µM IPTG. Cultures were diluted 1:4 in LB media containing 50 µg/ml of ampicillin and 70 µM IPTG (i) only or supplemented with indicated concentrations of (ii) SDS/EDTA or (iii) vancomycin. Cell growth was determined by measuring optical density at 600 nm at indicated time points.

Figure S6



Comparing σE Activities of WT and Δ*tig* Cells

σE activity was measured at various optical densities in both wild-type and Δ*tig::kan* cells by monitoring β-galactosidase expression from a reporter integrated into the chromosome consisting of the σE-dependent promoter *rpoHP3* fused to *lacZ*.

Figure S7

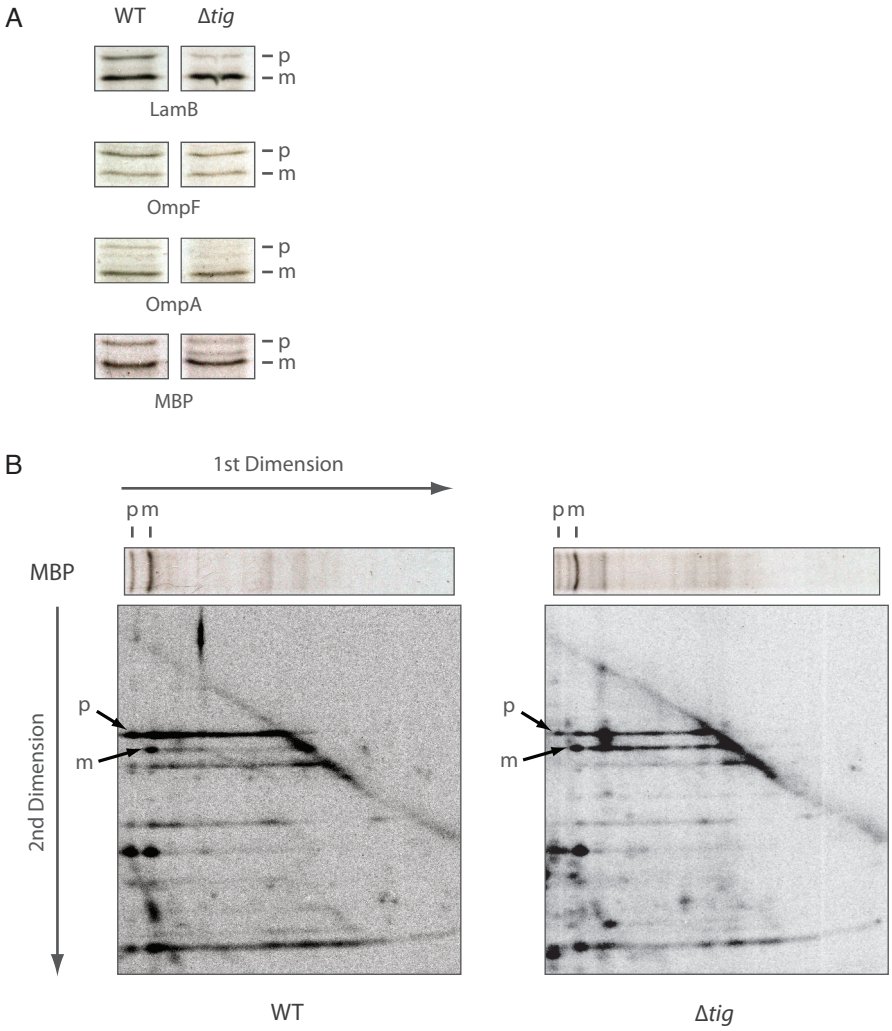

A
WT
Δtig
p
m
LamB
OmpF
OmpA
MBP
B
1st Dimension
p m
MBP
2nd Dimension
p
m
WT
Δtig

Figure S7

Cotranslational Translocation Occurs More Often for Sec-Dependent Substrates in Δ*tig* Cells

(A) Wild-type and Δ*tig::kan* cells grown in maltose minimal media were pulse-labeled with ^{35}S-methionine for 30 seconds. Labeling was stopped by adding 5% TCA. Cell extracts were immunoprecipitated using antisera specific for LamB, OmpF, OmpA, and MBP. Resulting pull-downs were resolved by 12% SDS-PAGE. Running positions of precursor (p) and mature (m) forms are marked for each protein. **(B)** 6-cm gel slices were excised from the first dimension (as in Figure S7A), digested with Staphylococcus aureus V8 protease, and resolved by 15% SDS-PAGE. Streaks emanating from precursor (p) or mature (m) MBP (indicated by arrows) respectively correspond to those containing or lacking their signal sequence.

REFERENCES

Ades, S.E., Connolly, L.E., Alba, B.M., and Gross, C.A. (1999). The Escherichia coli sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. Genes Dev. *13*, 2449–2461.

Agashe, V.R., Guha, S., Chang, H.-C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.U., and Barral, J.M. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. Cell *117*, 199–209.

Albanèse, V., Yam, A.Y.-W., Baughman, J., Parnot, C., and Frydman, J. (2006). Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. Cell *124*, 75–88.

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

Ball, L.A., and Kaesberg, P. (1973). Cleavage of the N-terminal formylmethionine residue from a bacteriophage coat protein in vitro. J. Mol. Biol. *79*, 531–537.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science *289*, 905–920.

Baram, D., Pyetan, E., Sittner, A., Auerbach-Nevo, T., Bashan, A., and Yonath, A. (2005). Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. Proc. Natl. Acad. Sci. USA *102*, 12017–12022.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*, 53–59.

Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolić, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B., and Ban, N. (2008). A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature *452*, 108–111.

Catrein, I., Herrmann, R., Bosserhoff, A., and Ruppert, T. (2005). Experimental proof for a signal peptidase I like activity in Mycoplasma pneumoniae, but absence of a gene encoding a conserved bacterial type I SPase. FEBS J. *272*, 2892–2900.

Chaba, R., Grigorova, I.L., Flynn, J.M., Baker, T.A., and Gross, C.A. (2007). Design principles of the proteolytic cascade governing the sigmaE-mediated envelope stress response in Escherichia coli: keys to graded, buffered, and rapid signal transduction. Genes Dev. *21*, 124–136.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805.

Crooke, E., and Wickner, W. (1987). Trigger factor: a soluble protein that folds pro-OmpA into a membrane-assembly-competent form. Proc. Natl. Acad. Sci. USA *84*, 5216–5220.

Datta, A.K., and Burma, D.P. (1972). Association of ribonuclease I with ribosomes and their subunits. J. Biol. Chem. *247*, 6795–6801.

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.

Eisner, G., Moser, M., Schäfer, U., Beck, K., and Müller, M. (2006). Alternate recruitment of signal recognition particle and trigger factor to the signal sequence of a growing nascent polypeptide. J. Biol. Chem. *281*, 7172–7179.

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., and Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature *431*, 590–596.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guthrie, B., and Wickner, W. (1990). Trigger factor depletion or overproduction causes defective cell division but does not block protein export. J. Bacteriol. *172*, 5555–5562.

Hagan, C.L., Silhavy, T.J., and Kahne, D.E. (2011). β-barrel membrane protein assembly by the Bam complex. Annu. Rev. Biochem. *80*, 189–210.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science *320*, 362–365.

Hoffmann, A., Merz, F., Rutkowska, A., Zachmann-Brand, B., Deuerling, E., and Bukau, B. (2006). Trigger factor forms a protective shield for nascent polypeptides at the ribosome. J. Biol. Chem. *281*, 6539–6545.

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. Biochim. Biophys. Acta *1803*, 650–661.

Homann, O.R., and Johnson, A.D. (2010). MochiView: versatile software for genome browsing and DNA motif analysis. BMC Biol. *8*, 49.

Huber, D., Rajagopalan, N., Preissler, S., Rocco, M.A., Merz, F., Kramer, G., and Bukau, B. (2011). SecA interacts with ribosomes in order to facilitate post-translational translocation in bacteria. Mol. Cell *41*, 343–353.

Hundley, H.A., Walter, W., Bairstow, S., and Craig, E.A. (2005). Human Mpp11 J protein: ribosome-tethered molecular chaperones are ubiquitous. Science *308*, 1032–1034.

Ingolia, N.T. (2010). Genome-wide translational profiling by ribosome footprinting. Methods Enzymol. *470*, 119–142.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Josefsson, L.G., and Randall, L.L. (1981). Different exported proteins in E. coli show differences in the temporal mode of processing in vivo. Cell *25*, 151–157.

Kadokura, H., and Beckwith, J. (2009). Detecting folding intermediates of a protein as it passes through the bacterial translocation channel. Cell *138*, 1164–1173.

Kaiser, C.M., Chang, H.-C., Agashe, V.R., Lakshmipathy, S.K., Etchells, S.A., Hayer-Hartl, M., Hartl, F.U., and Barral, J.M. (2006). Real-time observation of trigger factor function on translating ribosomes. Nature *444*, 455–460.

Keseler, I.M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R.P., Johnson, D.A., Krummenacker, M., Nolan, L.M., Paley, S., Paulsen, I.T., et al. (2009). EcoCyc: a comprehensive view of Escherichia coli biology. Nucleic Acids Res. *37*, D464–D470.

Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E., and Bukau, B. (2002). L23 protein functions as a chaperone docking site on the ribosome. Nature *419*, 171–174.

Kramer, G., Rutkowska, A., Wegrzyn, R.D., Patzelt, H., Kurz, T.A., Merz, F., Rauch, T., Vorderwülbecke, S., Deuerling, E., and Bukau, B. (2004). Functional dissection of Escherichia coli trigger factor: unraveling the function of individual domains. J. Bacteriol. *186*, 3777–3784.

Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat. Struct. Mol. Biol. *16*, 589–597.

Lee, H.C., and Bernstein, H.D. (2002). Trigger factor retards protein export in Escherichia coli. J. Biol. Chem. *277*, 43527–43535.

Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. Nature *466*, 1120–1124.

Martinez-Hackert, E., and Hendrickson, W.A. (2009). Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. Cell *138*, 923–934.

Mecsas, J., Rouviere, P.E., Erickson, J.W., Donohue, T.J., and Gross, C.A. (1993). The activity of sigma E, an Escherichia coli heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. Genes Dev. *7*(12B), 2618–2628.

Merz, F., Hoffmann, A., Rutkowska, A., Zachmann-Brand, B., Bukau, B., and Deuerling, E. (2006). The C-terminal domain of Escherichia coli trigger factor represents the central module of its chaperone activity. J. Biol. Chem. *281*, 31963–31971.

Merz, F., Boehringer, D., Schaffitzel, C., Preissler, S., Hoffmann, A., Maier, T., Rutkowska, A., Lozza, J., Ban, N., Bukau, B., and Deuerling, E. (2008). Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. EMBO J. *27*, 1622–1632.

Morris, D.R., and Geballe, A.P. (2000). Upstream open reading frames as regulators of mRNA translation. Mol. Cell. Biol. *20*, 8635–8642.

Müller, M., and Blobel, G. (1984). Protein export in Escherichia coli requires a soluble activity. Proc. Natl. Acad. Sci. USA *81*, 7737–7741.

Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., et al. (2011). Phenotypic landscape of a bacterial cell. Cell *144*, 143–156.

Patzelt, H., Kramer, G., Rauch, T., Schönfeld, H.-J., Bukau, B., and Deuerling, E. (2002). Three-state equilibrium of Escherichia coli trigger factor. Biol. Chem. *383*, 1611–1619.

Randall, L.L. (1983). Translocation of domains of nascent periplasmic proteins across the cytoplasmic membrane is independent of elongation. Cell *33*, 231–240.

Randall, L.L., and Hardy, S.J.S. (2002). SecB, one small chaperone in the complex milieu of the cell. Cell. Mol. Life Sci. *59*, 1617–1623.

Riggs, P.D., Derman, A.I., and Beckwith, J. (1988). A mutation affecting the regulation of a secA-lacZ fusion defines a new sec gene. Genetics *118*, 571–579.

Rutkowska, A., Mayer, M.P., Hoffmann, A., Merz, F., Zachmann-Brand, B., Schaffitzel, C., Ban, N., Deuerling, E., and Bukau, B. (2008). Dynamics of trigger factor interaction with translating ribosomes. J. Biol. Chem. *283*, 4124–4132.

Schlünzen, F., Wilson, D.N., Tian, P., Harms, J.M., McInnes, S.J., Hansen, H.A.S., Albrecht, R., Buerger, J., Wilbanks, S.M., and Fucini, P. (2005). The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. Structure *13*, 1685–1694.

Tenson, T., and Ehrenberg, M. (2002). Regulatory nascent peptides in the ribosomal tunnel. Cell *108*, 591–594.

Teter, S.A., Houry, W.A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C., and Hartl, F.U. (1999). Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell *97*, 755–765.

Tomic, S., Johnson, A.E., Hartl, F.U., and Etchells, S.A. (2006). Exploring the capacity of trigger factor to function as a shield for ribosome bound polypeptide chains. FEBS Lett. *580*, 72–76.

Ullers, R.S., Houben, E.N.G., Raine, A., ten Hagen-Jongman, C.M., Ehrenberg, M., Brunner, J., Oudega, B., Harms, N., and Luirink, J. (2003). Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the Escherichia coli ribosome. J. Cell Biol. *161*, 679–684.

Ullers, R.S., Houben, E.N.G., Brunner, J., Oudega, B., Harms, N., and Luirink, J. (2006). Sequence-specific interactions of nascent Escherichia coli polypeptides with trigger factor and signal recognition particle. J. Biol. Chem. *281*, 13999–14005.

Ullers, R.S., Ang, D., Schwager, F., Georgopoulos, C., and Genevaux, P. (2007). Trigger Factor can antagonize both SecB and DnaK/DnaJ chaperone functions in Escherichia coli. Proc. Natl. Acad. Sci. USA *104*, 3101–3106.

Vorderwülbecke, S., Kramer, G., Merz, F., Kurz, T.A., Rauch, T., Zachmann-Brand, B., Bukau, B., and Deuerling, E. (2004). Low temperature or GroEL/ES overproduction permits growth of Escherichia coli cells lacking trigger factor and DnaK. FEBS Lett. *559*, 181–187.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol. *16*, 274–280.

CHAPTER FIVE

Discussion

How can we best catalog the nascent proteome? In 1969, Joan Steitz showed that ribosomes can protect mRNA footprints when treated with ribonucleases (Steitz, 1969), thereby marking the position of translating ribosomes. Forty years later, sequencing technologies became powerful enough to enable the rapid identification of these ribosome-protected footprints (Bentley et al., 2008), giving birth to what is now called ribosome profiling (Ingolia et al., 2009). This approach has filled a critical gap in the systematic study of biological macromolecules. Given its ability to monitor translation in vivo with precision and sensitivity, use of ribosome profiling has spread rapidly across the scientific community (Bazzini et al., 2012; Brar et al., 2012; Guo et al., 2010; Hsieh et al., 2012; Ingolia et al., 2009; Ingolia et al., 2011; Li et al., 2012; Oh et al., 2011; Reid and Nicchitta, 2012; Stadler and Fire, 2011; Thoreen et al., 2012).

Bacterial Ribosome Profiling

Ribosome profiling of bacterial cells has provided much insight into prokaryotic translation (Li et al., 2012; Oh et al., 2011). Bacteria are fundamentally different than eukaryotes. First, bacteria do not compartmentalize their genomic DNA, so translation can begin as soon as the mRNA template is transcribed. Additionally, many bacterial transcripts are polycistronic, meaning that they encode for more than one protein on the same mRNA. Interestingly, translation rates of genes in the same operon poorly correlate with each other (Oh et al., 2011). Instead, expression levels seem to reflect the overall subunit composition of higher-order assemblies (unpublished data). For example, expression rates correspond well with the subunit composition of ATP synthase (F_OF_1 complex) whose nine subunits are encoded in the same operon. As a result, the α and β subunits of the F_1 subcomplex—each present in three copies—are expressed ~3-fold more than the γ, ∂, and ε subunits, all of which associate one copy. Excitingly, this result

suggests that subunit expression is regulated at the level of translation control, but more work is required to validate this concept.

What drives transient pausing of ribosomes in bacteria? The long-held assumption was that tRNA levels dictate the speed of ribosomes (Varenne et al., 1984). Yet our data did not support this hypothesis (Li et al., 2012). Instead, the anti-Shine-Dalgarno sequence of the 30S ribosomal subunit initiates short-lived interactions with Shine-Dalgarno-like sequences of open reading frames, ultimately driving transient arrest. Conversely, what can we learn about regulated ribosome pausing (Tenson and Ehrenberg, 2002)? By profiling strains that induce nascent chain-mediated translation arrest, we may better understand how this particular mechanism is exploited by cells.

Selective Ribosome Profiling

By profiling a specific population of ribosomes through the isolation of associated factors, one can investigate co-translational processes (Del Alamo et al., 2011; Oh et al., 2011). Moreover, we can better understand how these ribosome-associated factors cooperate with one another. For example, by selectively profiling bacterial ribosomes engaged by the peptidyl deformylase (Bingel-Erlenmeyer et al., 2008), methionine aminopeptidase (Ball and Kaesberg, 1973), SecA (Huber et al., 2011), and translocon (Mitra et al., 2005), we may gain insight into the ordered action of such processing enzymes and targeting factors, specifically in relation to the molecular chaperone trigger factor (TF) (Oh et al., 2011).

We do not fully comprehend the peptidyl prolyl *cis*/trans isomerase (PPIase) domain of TF (Hesterkamp et al., 1996; Stoller et al., 1995), because its deletion has no discernible effect in *E. coli* (Kramer et al., 2004). Intriguingly, mutations in RopA—the homologue of TF in the Gram-positive bacterium *S. pyogenes*—can prevent the export

of presecretory protein SpeB (Lyon et al., 1998), a cysteine protease that requires proline isomerization for function (Lyon and Caparon, 2003). Ultimately, proline isomerization by TF may not even occur cotranslationally. Interestingly, SurA chaperones β-barrel proteins in the periplasm (Hagan et al., 2011), and holds two PPIase domains (Merz et al., 2006), suggesting that TF may post-translationally isomerize prolines of such substrates. However, this model requires further validation.

We could extend this approach by enriching for ribosomes that have specific modifications or even those that are differentially localized (Reid and Nicchitta, 2012). For example, it may be possible to characterize the cotranslational status of disulfide bond formation, N-linked glycosylation, and ubiquitination-mediated degradation. The power of selective ribosome profiling relies on the ability to monitor the natural flux of substrates in wild-type cells. As a result, we anticipate that use of this approach will generate fascinating biological findings.

REFERENCES

Ball, L.A., and Kaesberg, P. (1973). Cleavage of the N-terminal formylmethionine residue from a bacteriophage coat protein in vitro. J. Mol. Biol. *79*, 531–537.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233–237.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R.*, et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*, 53–59.

Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolić, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B., and Ban, N. (2008). A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature *452*, 108–111.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science *335*, 552–557.

Del Alamo, M., Hogan, D.J., Pechmann, S., Albanese, V., Brown, P.O., and Frydman, J. (2011). Defining the Specificity of Cotranslationally Acting Chaperones by Systematic Analysis of mRNAs Associated with Ribosome-Nascent Chain Complexes. PLoS Biol. *9*, e1001100.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Hagan, C.L., Silhavy, T.J., and Kahne, D. (2011). β-Barrel membrane protein assembly by the Bam complex. Annu. Rev. Biochem. *80*, 189–210.

Hesterkamp, T., Hauser, S., Lütcke, H., and Bukau, B. (1996). Escherichia coli trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. Proc. Natl. Acad. Sci. USA *93*, 4437–4441.

Hsieh, A.C., Liu, Y., Edlind, M.P., Ingolia, N.T., Janes, M.R., Sher, A., Shi, E.Y., Stumpf, C.R., Christensen, C., Bonham, M.J.*, et al.* (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. Nature *485*, 55–61.

Huber, D., Rajagopalan, N., Preissler, S., Rocco, M.A., Merz, F., Kramer, G., and Bukau, B. (2011). SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria. Mol. Cell *41*, 343–353.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell *147*, 789–802.

Kramer, G., Rutkowska, A., Wegrzyn, R.D., Patzelt, H., Kurz, T.A., Merz, F., Rauch, T., Vorderwülbecke, S., Deuerling, E., and Bukau, B. (2004). Functional dissection of Escherichia coli trigger factor: unraveling the function of individual domains. J. Bacteriol. *186*, 3777–3784.

Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Lyon, W.R., and Caparon, M.G. (2003). Trigger factor-mediated prolyl isomerization influences maturation of the Streptococcus pyogenes cysteine protease. J. Bacteriol. *185*, 3661–3667.

Lyon, W.R., Gibson, C.M., and Caparon, M.G. (1998). A role for trigger factor and an rgg-like regulator in the transcription, secretion and processing of the cysteine proteinase of Streptococcus pyogenes. EMBO J. *17*, 6263–6275.

Merz, F., Hoffmann, A., Rutkowska, A., Zachmann-Brand, B., Bukau, B., and Deuerling, E. (2006). The C-terminal domain of Escherichia coli trigger factor represents the central module of its chaperone activity. J. Biol. Chem. *281*, 31963–31971.

Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C.L., Ban, N., and Frank, J. (2005). Structure of the E. coli protein-conducting channel bound to a translating ribosome. Nature *438*, 318–324.

Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G.*, et al.* (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. Cell *147*, 1295–1308.

Reid, D.W., and Nicchitta, C.V. (2012). Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. J. Biol. Chem. *287*, 5518–5527.

Stadler, M., and Fire, A. (2011). Wobble base-pairing slows in vivo translation elongation in metazoans. RNA *17*, 2063–2073.

Steitz, J.A. (1969). Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature *224*, 957–964.

Stoller, G., Rücknagel, K.P., Nierhaus, K.H., Schmid, F.X., Fischer, G., and Rahfeld, J.U. (1995). A ribosome-associated peptidyl-prolyl cis/trans isomerase identified as the trigger factor. EMBO J. *14*, 4939–4948.

Tenson, T., and Ehrenberg, M. (2002). Regulatory nascent peptides in the ribosomal tunnel. Cell *108*, 591–594.

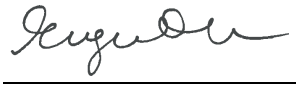
Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *485*, 109–113.

Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol. *180*, 549–576.

PUBLISHING AGREEMENT

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

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



Author Signature

7/26/12

Date