

A SUITE OF TOOLS FOR REPORTING AND
ENGINEERING PROTEIN FOLDING,
INTERACTIONS, AND POST-TRANSLATIONAL
MODIFICATIONS IN THE BACTERIAL
PERIPLASM

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Thomas J. Mansell

January 2012

© 2012 Thomas J. Mansell
ALL RIGHTS RESERVED

A SUITE OF TOOLS FOR REPORTING AND ENGINEERING PROTEIN
FOLDING, INTERACTIONS, AND POST-TRANSLATIONAL
MODIFICATIONS IN THE BACTERIAL PERIPLASM

Thomas J. Mansell, Ph.D.

Cornell University 2012

Therapeutic protein drugs are part of an emerging new generation of pharmaceutical products. However, production of such drugs is expensive due to the complex nature of many human proteins and post-translational modifications required for physiological efficacy and pharmacokinetic activity. We have developed several tools to improve production of human-like proteins in a relatively inexpensive host, *Escherichia coli*. First, we developed a system for monitoring and engineering protein solubility in the bacterial periplasm, a compartment with many useful features for heterologous protein expression. Next, we developed a system for monitoring protein-protein interactions *in vivo* in the periplasm, which we can leverage for the production of novel and improved antibodies and antibody fragments. Building upon our protein folding reporter technology, we have developed a system for examining the effect of N-linked glycosylation, an important post-translational modification, on protein folding *in vivo*. This system allows us to (1) study the effect of glycosylation on folding of various glycoproteins from pathogenic organisms and (2) create a genetic selection with the purpose of engineering the glycosylation pathway using the versatile *E. coli* as a host. Finally, we have created a modified genome-scale flux balance analysis model of *E. coli* to determine *in silico* metabolic factors that affect glycosylation efficiency.

BIOGRAPHICAL SKETCH

Tom Mansell was born in Hershey, Pennsylvania on January 10, 1983 during a snowstorm. He was the valedictorian of the Class of 2000 at Blue Mountain High School in Schuylkill Haven, PA.

He received his Bachelor of Science degree in Chemical Engineering (with second major in Spanish) in May 2004 and Masters of Science in Engineering in Chemical Engineering in August 2005, both from The Johns Hopkins University. He was advised by Prof. Marc Ostermeier and pursued research on allosteric molecular switches. While at Hopkins, he sang with the painfully sexy Mental Notes *a cappella* group.

He began his PhD research at Cornell University in August 2005. Co-advised by Profs. Matthew DeLisa and Jeffrey Varner, he focused on developing selections for understanding and engineering protein folding and glycosylation in the periplasm of *E. coli*. In 2009, he was awarded the Corning, Inc. Foundation Graduate Science Fellowship. While at Cornell, he sang with the Cornell Glee Club and performed in several shows in the Department of Theatre, Film, and Dance and Risley Theatre.

After his Ph.D., Tom will pursue a postdoctoral research position at the University of Colorado at Boulder in the laboratory of Prof. Ryan Gill, where he will be designing orthogonal genetic circuit elements with broad applications to synthetic biology.

He was married to Sarah Goonan in May 2011 and enjoys writing about the science of wine and beer on *Palate Press: the Online Wine Magazine* and *The New York Cork Report*.

For Sarah, who worked just as hard as I did (maybe harder) towards this Ph.D.

ACKNOWLEDGEMENTS

I would first like to acknowledge the contributions of Prof. Matthew DeLisa to this work. Matt was surprisingly willing to branch out into the uncharted territory of engineering bacterial protein glycosylation when we first met back in 2005. Since then, he has been a constant fountain of ideas and insights along the somewhat bumpy road to biotechnological applications of glycoengineering in bacteria. I am lucky to count him as an advisor, mentor, and friend.

I am also grateful for the influence and guidance of Prof. Jeffrey Varner. Jeff has been an excellent mentor and experience has served those in his group well. In particular, I thank Jeff for exposing me, through group meetings and collaborative projects, to the importance and power of modeling in biological systems.

My third committee member, Prof. Michael Shuler, also provided interesting insights along the way throughout my Cornell career. He also helped lay the groundwork for my studies in both modeling and experimental work with his Bioprocess Engineering class, CHEME543.

I must also acknowledge the feedback, helpful discussions (scientific and otherwise), and overall contributions of past and present members of the DeLisa and Varner research groups. While I have gained valuable insights from almost everyone in these groups, I especially thank Dr. Adam Fisher, Dr. Rob Conrado, Dr. Amy Karlsson, Dr. Ryan Tasseff, Jason Boock, Alyse Portnoff, Juan Valderrama, and Cassie Guarino for their support and feedback along the way. Many experiments in this work were supported by undergraduate and graduate students that I have mentored over the years. Steve Linderman, Caroline Tabler, and Siddharth Ramshankar were outstanding undergraduates worthy of recognition.

I would also like to acknowledge the National Science Foundation, NYSTAR, and especially the Corning, Inc. Foundation for financial support of this work.

My family and especially my parents, Tom and Terry Mansell, have been a constant stream of support during my entire academic career and explaining my research to them along the way, while challenging, has proven to be a valuable exercise in communication, an often overlooked skill among academics and engineers. Finally, I thank my wife, Sarah Goonan Mansell for her tremendous support throughout the long process of the Ph.D. Whether helping me talk out a problem, distracting me from worrisome issues, providing comfort, or just preparing a hot meal, Sarah has been an essential part of my graduate study and will continue to be an essential part of my life.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vii
List of Figures	viii
1 Engineering the Protein Folding Landscape	1
2 A rapid protein folding assay the bacterial periplasm	32
3 An <i>in vivo</i> assay for bacterial <i>N-linked</i> glycosylation	63
4 Modeling bacterial <i>N-linked</i> glycosylation using an adapted genome-scale model of <i>Escherichia coli</i>	81
5 An Assay for Reporting Protein-Protein Interactions in the Bacterial Periplasm	95
6 Conclusions and Future Directions	104
Bibliography	111

LIST OF FIGURES

1.1	Strategies for engineering the protein folding landscape of <i>E. coli</i> . From its birth in the ribosome as a nascent polypeptide to its eventual attainment of a properly folded 3D conformation, a protein interacts with many components of the cellular biomachinery. Indicated are a number protein machinery targets that researchers have taken or are taking steps to engineer as a means to alter the cellular folding landscape.	4
1.2	The ribosomal exit tunnel. This cross section of the 50S subunit of the <i>E. coli</i> ribosome depicts the peptidyltransferase center (PTC, black with star), the emerging polypeptide (black line), and notable proteins bordering the exit tunnel (white). The bulk of the subunit is represented in grey. Peptide bond formation occurs in the PTC and the polypeptide must pass through the narrow (10-20 in diameter), long (100 in length) exit tunnel before emerging from the ribosome. Proteins L23 and L29 are believed to interact with the nascent peptide and moderate interactions with chaperones and other cytosolic proteins. Figure adapted from Jenni and Ban [51].	7
1.3	Schematic of the directed evolution process. By following an iterative cycle of DNA mutagenesis to introduce diversity and selection to identify unique clones with the desired phenotype, researchers can obtain proteins (or protein complexes) with improved or even vastly different functions.	14
2.1	Schematic of SRP-mediated periplasmic folding reporter. (a) The DsbA signal peptide enables co-translational translocation across the cytoplasmic membrane via the SRP-dependent pathway. Nascent polypeptides enter the periplasm where they fold properly, conferring Amp resistance to cells, or misfold, leading to eventual aggregation and/or proteolysis. (b) Fusion proteins created in this study contained an N-terminal DsbA signal peptide (ssDsbA) followed by a POI and a C-terminal selectable marker (TEM-1 Bla). Short linker sequences (SR, VDGS) were created at the junctions.	35

- 2.2 Redox-dependent folding reported via SRP selection. (a) An equivalent number of DHB4 (*dsbA*⁺) and DHA (*dsbA*⁻) cells expressing either ssDsbA-PhoA-Bla, ssTorA-PhoA-Bla, ssDsbA-Bla or ssDsbA-scFv13-Bla were spot-plated on LB/agar containing either 50 $\mu\text{g/ml}$ Cm (- Amp) or 100 $\mu\text{g/ml}$ Amp (+ Amp). (b) Western blot analysis of periplasmic (per) and cytoplasmic (cyt) fractions derived from DHB4 (*dsbA*⁺) and DHA (*dsbA*⁻) cells expressing either ssDsbA-PhoA-Bla, ssTorA-PhoA-Bla or ssDsbA-scFv13-Bla. Samples were blotted with Bla-specific antibodies. 38
- 2.3 A universal folding reporter. (a) Western blot analysis of whole cell lysates generated from an equivalent number of DH5 α cells expressing the ssDsbA-POI-Bla constructs as indicated. Samples were blotted with Bla-specific antibodies. (b) An equivalent number of DH5 α cells expressing ssDsbA-MetF-Bla or ssDsbA-MetK-Bla were spot-plated on LB/agar containing either 50 $\mu\text{g/ml}$ Cm (- Amp) or 100 $\mu\text{g/ml}$ Amp (+ Amp). Nitrocefin hydrolysis activity of whole-cell lysates from cells expressing MetF-Bla and MetK-Bla with or without the ssDsbA signal peptide. Activity was measured as the initial velocity of absorbance change at 486 nm. Relative Bla activity was obtained by normalizing to the activity of the fusion without the signal peptide. Bla activity was measured in triplicate and error bars represent standard error of the mean. (c) DH5 α cells expressing ssDsbA-GFP-Bla or ssTorA-GFP-Bla were spot-plated as in (b). Fluorescence microscopy analysis (100X) of DH5 α cells expressing ssDsbA-GFP-Bla or ssTorA-GFP-Bla as indicated. (d) Western blot analysis of periplasmic (per) and cytoplasmic (cyt) fractions generated from an equivalent number of DH5 α cells expressing ssDsbA-POI-FLAG fusions lacking the Bla moiety for the clones indicated. Blots were probed by anti-FLAG antibodies. GroEL was used as a fractionation marker by probing with anti-GroEL antibodies (data not shown). GRB2 per and RAF1 per samples were included as positive loading controls for the blot on the right. MBC values ($\mu\text{g/ml}$ Amp) corresponding to each protein are indicated below the panels 45

- 2.4 Effect of cis- and trans-acting factors on periplasmic protein folding. (a) DH5 α cells expressing wild-type (wt) MBP and its variants MalE31, I33P, G32D as sandwich fusions between ssDsbA and Bla. An equivalent number of cells were spot-plated on 100 $\mu\text{g}/\text{ml}$ Amp (+ Amp) or 50 $\mu\text{g}/\text{ml}$ Cm (- Amp). Open triangle indicates increasing solubility of MBP. (b) An equivalent number of DH5 α cells co-expressing periplasmic chaperones FkpA, Skp, and SurA (or with empty pBAD18-Kan) along with wild-type (wt) MBP or MalE31 fusions were spot plated at various dilution factors (as indicated to left) on LB/agar supplemented with 200 $\mu\text{g}/\text{ml}$ Amp and 50 $\mu\text{g}/\text{ml}$ Kan. (c) Western blot analysis of periplasmic fractions from cells co-expressing FkpA, Skp, and SurA chaperones as indicated along with ssDsbA-MalE31-Bla. Samples were probed with Bla-specific antibodies. Detection of DsbA using anti-DsbA serum was performed as a loading control. (d) An equivalent number of cells co-expressing FkpA, Skp, and SurA chaperones as indicated along with ssDsbA-Bla were spot-plated on 50 $\mu\text{g}/\text{ml}$ Kan and either 800 $\mu\text{g}/\text{ml}$ Amp (+ Amp) or 20 $\mu\text{g}/\text{ml}$ Cm (- Amp). 50
- 2.5 Directed evolution of the A β 42 peptide. (a) Spot plating of an equivalent number of cells expressing fusions containing wild-type A β 42 and solubility-enhanced variants GM11, GM7 and GM6 on 100 $\mu\text{g}/\text{ml}$ Amp (+ Amp) or 50 $\mu\text{g}/\text{ml}$ Cm (- Amp). The variants are derivatives of A β 42 with the following mutations: GM11 (H6Q/V12A/V24A/132M/V36G), GM7 (V12A/I32T/L34P) and GM6 (F19S/L34P). (b) Spot plating of an equivalent number cells expressing evolved variants of A β 42 with greater solubility than wild-type A β 42 at various dilution factors from overnight cultures. For comparison, wild-type A β 42 and soluble variant GM6 are shown. (c) Amino acid sequences of evolved A β 42 peptides, along with A β 42 and GM6 for comparison. 54

3.1	Glycosylation-dependent folding reported via SRP selection. (a) An equivalent number of CLM24 cells expressing ssDsbA-Peb3-Bla (WT or K135E) and either pACYC-pglKan (pgl+) or pACYC-pglmutKan (pgl-) were spot-plated on LB/agar containing 200 $\mu\text{g}/\text{ml}$ Amp and 50 $\mu\text{g}/\text{ml}$ Kan. (b) Western blot analysis of periplasmic fractions derived from CLM24 cells expressing the above constructs. Samples were blotted with Bla-specific antibodies (anti-BLA) or hR6 serum, which is specific to the <i>C. jejuni</i> heptasaccharide. (c) Nitrocefin hydrolysis activity of periplasmic fractions of above constructs. Activity was measured as the initial velocity of absorbance change at 486 nm. Specific Bla activity was obtained by normalizing to the total protein present in the periplasmic fraction. Bla activity was measured in triplicate and error bars represent standard error of the mean.	69
3.2	Glycosylation-dependent folding of CjaA and Cj0610c. (a) Spot plating as in Figure 3 on 200 $\mu\text{g}/\text{mL}$ Amp and 50 $\mu\text{g}/\text{mL}$ Kan (b) Western blot analysis as in Figure 3 (c) Nitrocefin hydrolysis activity as in Figure 3	72
3.3	Glycosylation-dependent folding of im7 wild-type, DQNAT, and ENNAT constructs (a) Spot plating as in Figure 3 on 400 $\mu\text{g}/\text{mL}$ Amp and 50 $\mu\text{g}/\text{mL}$ Kan (b) Western blot analysis as in 3 (c) Nitrocefin hydrolysis activity as in 3	75
4.1	Glycosylation in <i>C. jejuni</i> and <i>E. coli</i> . Constituent nucleotide sugars (GalNAc, blue square; glucose, blue circle; GlcNAc, blue square; bacillosamine, orange star) are synthesized in the cytoplasm and attached to a membrane-bound lipid carrier, undecaprenol (UDCP). The glycan is then transported to the periplasmic side of the inner membrane, where it is transferred to the asparagine residue in a protein containing the acceptor site. . . .	83
4.2	Reactions of the central carbon metabolism that are coupled to protein glycosylation. Arrows in bold indicate coupled reactions.	87
4.3	Reactions of the central carbon metabolism that are coupled to protein glycosylation. Arrows in bold indicate coupled reactions.	91
5.1	(a) Schematic of the plasmids used in this work. The reporter is a two-plasmid system in which genes of interest are fused to the inactive fragments of BLA. Fos and Jun are used here as examples, and proteins of interest would replace Fos and Jun in their respective constructs. (b) Schematic of the protein complementation assay. Protein interactions result in formation of an active BLA complex which is able to hydrolyze ampicillin, conferring an antibiotic-resistance phenotype.	97

5.2	Protein complementation of interacting domains. (a) Spot plating of Fos and Jun leucine zippers on 50 $\mu\text{g}/\text{ml}$ Amp. GCN4, another leucine zipper, but one that does not interact with Fos, is used as a negative control. (b) Spot plating as in (a) of antibody fragment scFv-GCN4 and its antigen, the GCN4 leucine zipper. GCN4-PP is a double point mutant to wild-type GCN4 which reduces homodimerization of the leucine zippers. Jun is used as a negative control. (c) Spot plating as in (a) of scFv D10[124], with affinity to phage protein Gpd. Jun and GCN4 used as negative controls.	100
5.3	Interaction of Fos and Jun variants (a) Spot plating as in Figure 5.2 of Fos/Jun variants. Note: notations such as L3V refer replacement of leucine 3 with valine, not necessarily residue 3. Similar growth for all variants was confirmed on Cm/Kan plates (data not shown) (b) MIC (blue) and MBC (red) measurements of Fos and Jun variants pictured in (a).	101
6.1	Analysis of RNase A-BLA fusions. (a) Spot plating as in Figure 3 on 200 $\mu\text{g}/\text{mL}$ Amp and 50 $\mu\text{g}/\text{mL}$ Kan (b) Western blot analysis as in Figure 3 (c) Nitrocefin hydrolysis activity as in Figure 3. (d) Western blotting with hR6 of RNaseA (no BLA fusion) fused to N-terminal signal sequences of DsbA, MalE, and PhoA.	106
6.2	Schematic of the protein-carbohydrate interaction complementation assay. Binding of a sugar group by a carbohydrate binding protein results in formation of an active BLA complex which is able to hydrolyze ampicillin, conferring an antibiotic-resistance phenotype.	109

CHAPTER 1

ENGINEERING THE PROTEIN FOLDING LANDSCAPE

Protein folding¹ - the attainment of a discrete, stable, three-dimensional structure - is the primary prerequisite for protein function. Since proteins gradually evolve over long times with changing functions and structures, it is reasonable to assume that protein folding presents a significant obstacle for living systems. Moreover, the ability of microorganisms such as bacteria to reliably produce polypeptide chains that fold into functional proteins is a cornerstone of biotechnology. Thus, the inability to fold recombinant proteins *in vivo* represents a major bottleneck of the biotechnology enterprise [11]. To remedy this situation will ultimately require an improved understanding of the mechanisms of protein folding and aggregation in the complex environment of the cell. This is a formidable challenge because much of what we know regarding the protein folding process comes from *in vitro* studies using a wide array of sophisticated approaches that are difficult or impossible to implement *in vivo*. Furthermore, *in vitro* experiments may be of limited value in predicting *in vivo* folding behavior for the same protein as it is now well established that protein folding under physiological conditions in the presence of all other cellular constituents differs greatly from protein folding in a test tube of aqueous buffer [59, 101]. In the intracellular environment, protein folding and solubility are governed by the thermodynamics of folding in high concentrations of macromolecules, interaction with folding chaperones and isomerases, susceptibility to protease degradation, compatibility with secretory trafficking, as well as the general impact of folded protein expression on the fitness of cells. For instance, while some

¹Adapted with permission from Mansell TJ, Fisher AC, and DeLisa MP, "Engineering the Protein Folding Landscape" (2008) *Current Protein and Peptide Science*, 9 2, pp. 138-149

proteins are able to fold spontaneously *in vitro* without assistance from or interaction with other proteins, numerous others require interactions with accessory molecules like molecular chaperones To attain a properly folded conformation [89, 90, 99, 115, 255]. Still other proteins interact with chaperones To remain in an unfolded state, especially those proteins which will be transported via cellular translocation pathways [70]. Despite an incomplete understanding of the complexities of intracellular protein folding, a number of recent experimental advances have provided researchers with an unprecedented opportunity to (i) visualize proteins in their native environment [237] and (ii) directly enhance the folding properties of any particular protein. In the former case, imaging protein dynamics in living cells has been facilitated largely through the advent of genetically-encoded fluorophores such as the green fluorescent protein (GFP) although a number of next generation approaches have been developed recently that correlate a protein's dynamics with its changing structural state or activity [38]. In the latter case, modulating the folding behavior of a target protein sequence has been accomplished by making mutations either rationally or randomly to the primary structure of a protein and assaying for improved folding and/or function in living cells [151, 152, 241]. The most common targets for such a strategy are complex heterologous proteins that are often poorly expressed in simple host organisms (e.g., expression of mammalian proteins in *Escherichia coli*) [56]. However, for many proteins, especially therapeutic proteins, it is typically undesirable to alter the amino acid sequence even if it results in an improved protein as this may have severe consequences such as loss of efficacy or introduction of unexpected antigenicity. A more attractive approach is to move away from engineering proteins one-at-a-time, and instead to exploit our ever-increasing awareness of the factors influencing protein folding in bac-

teria [80, 10]. In this article, we review a number of interrelated efforts and the methods used therein to engineer better-folding cells by re-designing the intracellular landscape in a manner that universally improves the efficiency of the folding process (see Fig.1.1).

Engineering ribosomal translation

The central dogma of molecular biology, first enunciated by Francis Crick in 1958 [47] and reiterated in 1970 [48], is represented by three major stages: replication of DNA, transcription of DNA into mRNA and translation of mRNA into protein. The translation process is catalyzed by the ribosome, a 2.3 MDa-complex assembly of proteins and RNA that serves as the birthplace of proteins in the cell. Structural analysis of the entire ribosome in complex with bound tRNAs, mRNAs and chaperones [35, 68, 125, 205, 256] has resulted in a quantum jump in our understanding of how the genetic code is translated into proteins. Results of these and related studies have illuminated surprising aspects of cellular protein biosynthesis, including the active participation of the ribosome in the earliest stages of protein folding [45, 68, 96, 97, 123, 133, 147, 259]. Based on its ability to impact the protein folding process, the ribosome and its co-factors (tRNA, mRNA, ribosome-associated chaperones) are a prime target for researchers seeking to develop strategies that improve the folding efficiency of target proteins. Folding and translation rate. In living cells, there is a correlation between the frequency of codon usage and the level of cognate tRNA. As a result, translational problems such as frameshifting [195, 214], hopping [110], or premature termination of translation [79, 84, 91, 195] are often encountered when an mRNA contains high levels of rarely used codons (e.g., the AGA and

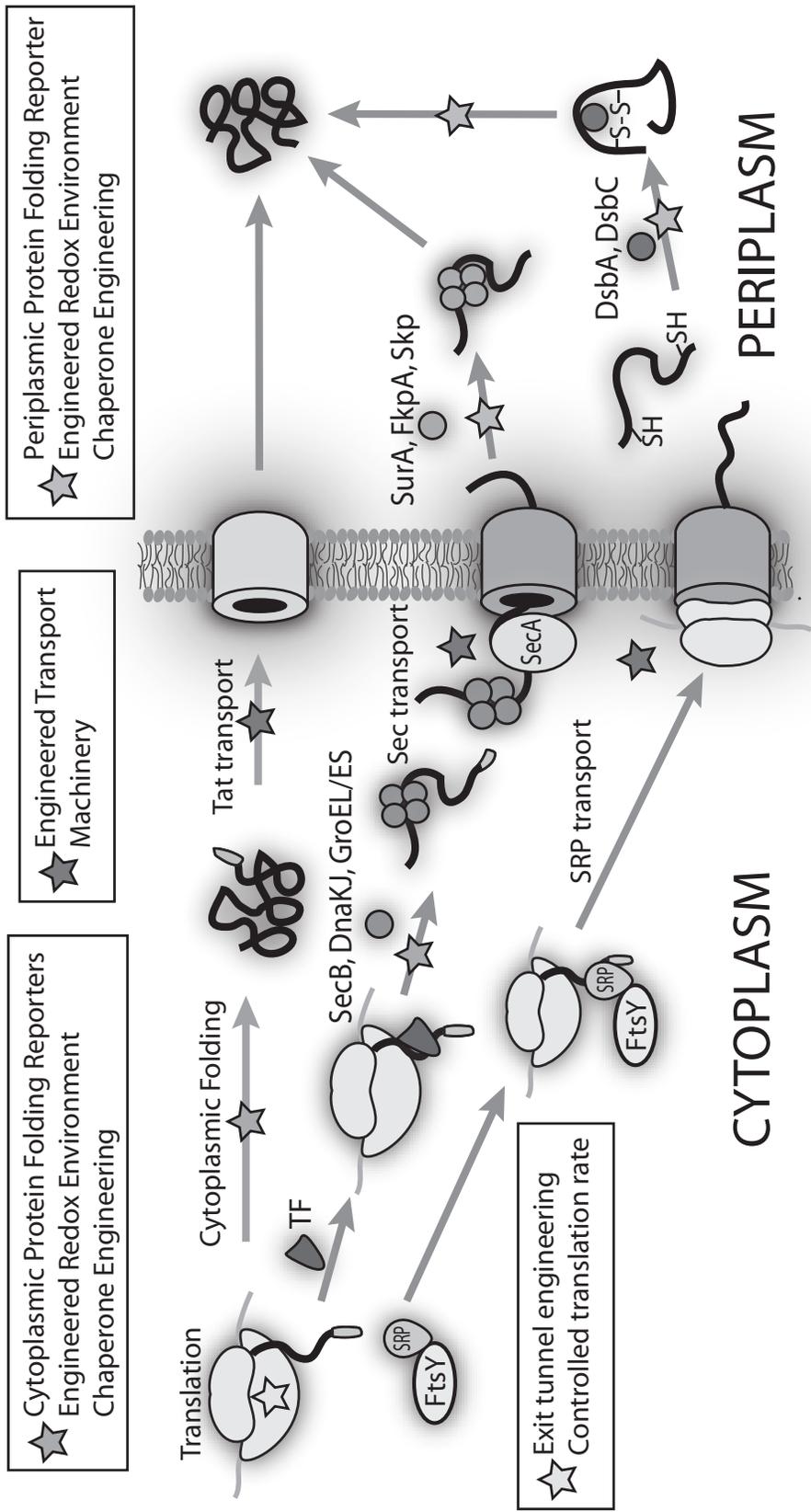


Figure 1.1: Strategies for engineering the protein folding landscape of *E. coli*. From its birth in the ribosome as a nascent polypeptide to its eventual attainment of a properly folded 3D conformation, a protein interacts with many components of the cellular biomachinery. Indicated are a number of protein machinery targets that researchers have taken or are taking steps to engineer as a means to alter the cellular folding landscape.

AGG codons for arginine are the least used codons in *E. coli* [169]). Due to the inherent relationship between translation rate and the folding of the nascent polypeptide chain, rare codons can be particularly aggravating for biotechnological applications where heterologous genes are expressed in a host organism whose codon bias may be incompatible with the sequence of the cloned gene [50, 109]. This problem can often be remedied via overexpression of low-abundance tRNAs which recognize rare codons (e.g., the Rosetta strain, see Table 1.1) [8] or via codon optimization where rare codons are replaced with synonymous codons that are preferred by the host organism [117, 127]. Indeed, rare codons can impart dramatic changes in the timing of translation: a 6- to 9-fold reduction in translation rate has been observed for sequences encoded with rare codons versus the preferred codons [20, 213]. Certain organisms have exploited this phenomenon to productively slow the kinetics of translation as a means of influencing the folding behavior of a growing nascent polypeptide chain. For instance, in the structural organization of complex multidomain proteins in *E. coli*, protein domain boundaries are largely coded by translationally slow regions on the RNA [225]. This suggests that induction of ribosome slowing or pausing may provide a mechanism for efficient co-translational folding of individual domains and reveals an intriguing mechanism of *in vivo* protein folding. However, care must be taken when altering translation rates as the introduction of certain synonymous codons can have a profound effect on the folding behavior of both endogenous [122] and heterologous [46] proteins. One study even demonstrated [122], that the presence of a rare synonymous polymorphism in human P-glycoprotein (P-gp) altered its conformation and substrate specificity, presumably by a change in the timing of co-translational folding on the ribosome. Based on the above findings, we speculate that while controlling the

speed of the ribosome via codon manipulation may cause folding problems in some cases, it could in other cases prove to be an effective strategy for synchronizing translation kinetics with the optimal folding rate of a given target protein. One intriguing candidate for ribosomal pausing is the C-terminus of the normally periplasmic SecM protein that encodes an arrest sequence that causes elongation cessation within the ribosome [232]. Introduction of this stall sequence at the C-terminus of a variety of recombinant proteins (e.g., GFP, single-chain antibody fragments) did not result in a loss of function but rather efficient stalling and display of functional proteins on ribosomes *in vivo* [44, 63].

Ribosomal exit tunnel. One potential region of the ribosome itself that may have the greatest influence on the early stages of protein folding is the protein- and RNA-lined exit tunnel, depicted in Fig. 1.2. This physical tunnel is approximately 100 Å long and ranges from 10-20 Å in diameter (large enough to accommodate an α -helix) and has been hypothesized to affect folding of nascent peptides *in vivo* [132, 105]. In support of this hypothesis, Johnson and coworkers reported that α -helix formation in transmembrane proteins was induced and stabilized by the ribosome exit tunnel [252]. Further evidence comes from molecular dynamics simulations demonstrating that translocation of a polypeptide through an exit tunnel-like channel can affect both the kinetics and thermodynamics of nascent polypeptide folding [45, 123, 259]. As our understanding of the exit tunnels role in protein folding increases, it is expected that intentional alterations to the exit tunnel proteins and/or RNA might yield substrate-optimized tunnels that favor *in vivo* folding of heterologous proteins. Along these lines, we have employed a directed evolution strategy (see Fig. (1.3)) to isolate variants of the exit tunnel protein L29 that may enhance the folding of GFP (Contreras-Martinez and DeLisa, unpublished observations).

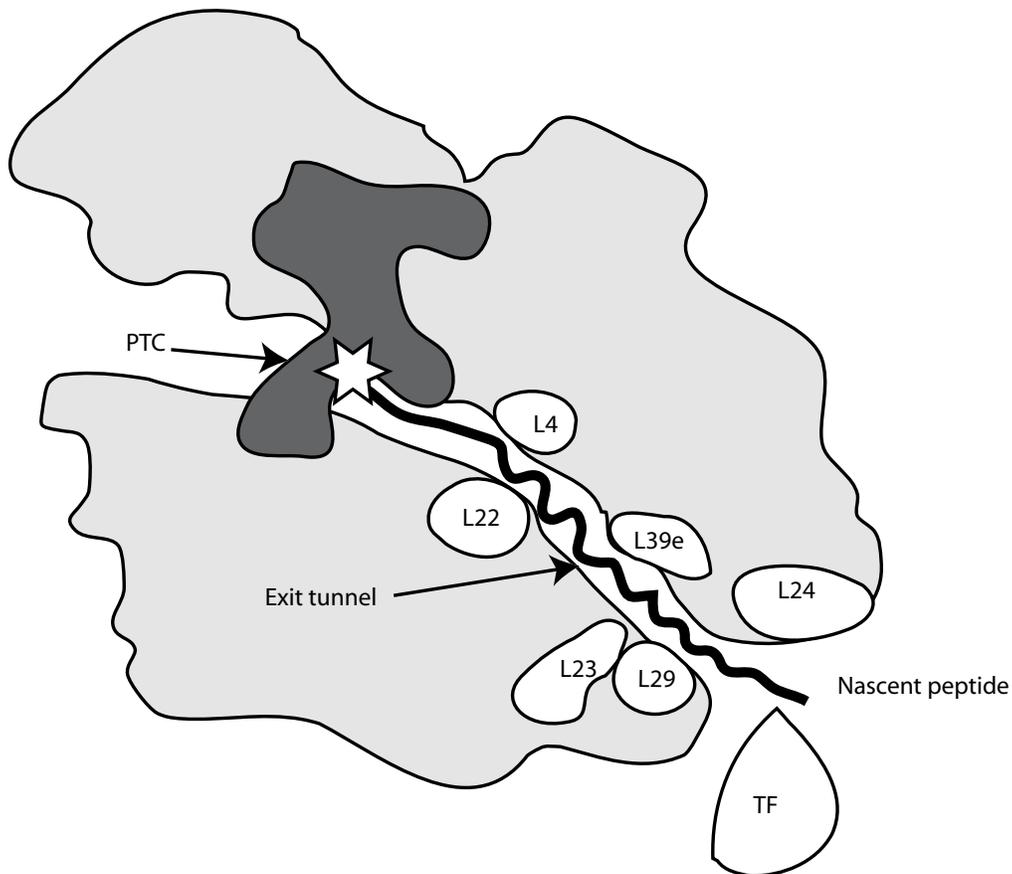


Figure 1.2: The ribosomal exit tunnel. This cross section of the 50S subunit of the *E. coli* ribosome depicts the peptidyltransferase center (PTC, black with star), the emerging polypeptide (black line), and notable proteins bordering the exit tunnel (white). The bulk of the subunit is represented in grey. Peptide bond formation occurs in the PTC and the polypeptide must pass through the narrow (10-20 in diameter), long (100 in length) exit tunnel before emerging from the ribosome. Proteins L23 and L29 are believed to interact with the nascent peptide and moderate interactions with chaperones and other cytosolic proteins. Figure adapted from Jenni and Ban [51].

Besides its direct role in protein folding, the ribosomal exit tunnel also serves as a docking site for molecular chaperones that engage the nascent polypeptide as it emerges from the tunnel. Most notably, the chaperone trigger factor (TF) is known to bind to ribosomes via contacts with the exit tunnel protein L23 and, to a lesser extent, L29 [12, 68, 133]. From its perch on the ribosome, TF creates a protected folding environment where nascent proteins may be shielded from proteases and aggregation [68]. Accordingly, overexpression of TF in the cytoplasm of *E. coli* with aggregation-prone recombinant proteins, such as human lysozyme or mouse endostatin, had marked effects on the soluble production of these proteins [173]. Like TF, the targeting factor signal recognition particle (SRP), which interacts specifically with the signal sequence in nascent inner membrane and periplasmic proteins, was found to interact extensively with the L23 and L29 proteins [232]. Since many membrane proteins are secreted via the SRP-dependent secretory pathway, we speculate that, by analogy to TF, SRP overexpression is likely to be useful for more efficient expression of heterologous inner membrane proteins such as eukaryotic G protein-coupled receptors (GPCRs), a class of proteins whose crystal structures remain elusive due to expression bottlenecks [158]. More recently, covalent attachment of recombinant proteins to exit tunnel components such as L23 and TF has yielded several new technologies including (i) a platform for soluble expression and purification of aggregation-prone proteins [212] and (ii) targeted proteolysis for conditional inactivation of essential proteins [95]. In addition to TF and SRP binding, the exit tunnel also plays a key role in binding the SecM stall sequence under secretion-defective conditions [170]. Specifically, the SecM arrest sequence interacts with the ribosomal exit tunnel via contacts with 23S rRNA and ribosomal protein L22 and causes elongation arrest within the ribosome [232]. This mechanism

was recently exploited for the development of an intracellular ribosome display technology that facilitated directed evolution of antibodies with enhanced cytosolic stability [44].

Engineering protein folding in the cytoplasm

Macromolecular crowding

After emerging from the ribosome, the polypeptide enters an environment crowded with macromolecules of protein, DNA, and RNA. One of the largest differences between protein folding in aqueous buffer and folding *in vivo* is this high concentration of macromolecules within the cell that can reach 300-400 g/L depending on cell cycle [58, 149]. While this macromolecular crowding is believed to generally contribute to *in vivo* protein folding by facilitating greater association of polypeptide chains with chaperones and thereby enhancing the activity of molecular chaperones [58] distributed throughout the cytoplasm, it can also be a barrier for correct folding, especially for proteins that are prone to aggregation [219]. Without chaperones, folding is physically more challenging in a crowded milieu although there are some proteins (e.g. apocytochrome b5) that prefer to fold when crowding agents are added to dilute solutions *in vitro*, and when expressed *in vivo* [29]. Since the folding phenomenon in the face of macromolecular crowding is inconsistent, a not-so-obvious target for modifying the protein folding landscape is the intracellular macromolecule concentration and composition. One intriguing (and ostensibly counterintuitive) way to modify the macromolecular concentration in bacteria for improved protein folding is the expression of the endoribonuclease toxin MazF that specifically

cleaves mRNAs at ACA sequences [257, 258]. The expression of this toxin leads to quasi-dormant cells that, in spite of complete cell growth inhibition, are capable of energy metabolism and the biosynthesis of RNA and protein [218]. These quasi-dormant cells provide a means of producing only a protein of interest, in high-yield, in the virtual absence of background protein expression. The major success of this feat of engineering was, following nucleotide optimization and the noncoding mutation of all ACA sequences, the production of soluble human eotaxin with greater than 90% purity. Intrinsic factors affecting cytoplasmic protein folding. A proteins structure is key to its function and its primary amino acid sequence is, of course, key to its structure. Changes in amino acid sequence can affect the function and folding properties of any given protein. With this in mind, protein engineers have often sought to enhance a proteins folding properties by altering its primary sequence using, for instance, (i) rational design to create more soluble peptides [194] and stability-enhanced serine proteases like subtilisin [28] and (ii) directed evolution (see Fig. (3)) to select for super-soluble or superfolder proteins [178]. However, a detailed discussion of how sequence-specific changes can be implemented to yield better folding proteins is outside the scope of this review and, instead the reader is referred to the following reviews [151, 152, 241].

Extrinsic factors affecting cytoplasmic protein folding

In addition to engineering of the protein itself, many steps have been taken to engineer the local environment in which protein folding occurs. For instance, numerous studies have shown that optimal production of certain proteins is temperature dependent [37], with many proteins responding favorably to ex-

pression at 25°C or 30°C. Even lower temperatures have proven to be beneficial for the expression of toxic and/or proteolytically sensitive proteins [166]. Two factors that may contribute to increased expression at lower temperatures are slower protein aggregation kinetics and diminished protease activity. Along similar lines, Gill et al. demonstrated that chemically-induced cell conditioning was an effective means of dynamically altering the stress gene response in a manner that was favorable for recombinant protein expression [81]. However, engineering the folding environment *in vivo* is not limited to changing extracellular conditions. For example, alteration of a proteins immediate environment via fusion to a well-expressed partner like *E. coli* maltose binding protein [196], thioredoxin [138], or glutathione-S-transferase [209] can dramatically increase expression of the target protein. Alternatively, genetic engineering has been used to re-design the cytoplasmic environment in a manner that promotes disulfide bond formation. In the cytoplasm of *E. coli*, the formation of disulfide bonds is strongly disfavored by the highly reducing environment of the cytoplasm, which is maintained by the proteins thioredoxin and glutaredoxin [186]. Since the stability of many heterologous proteins, particularly those of mammalian origin having multiple cysteine residues, depends on disulfide bond formation [219], cytoplasmic production of these proteins is generally confronted by misfolding, aggregation and inclusion body formation. To remedy this situation, Beckwith and colleagues used a screen based on the activity of an export-defective variant of the enzyme alkaline phosphatase (AP) to isolate a strain of *E. coli* in which stable disulfide bonds formed in the cytoplasm [54]. Since AP requires the formation of disulfide bonds to function, mutants were selected that recovered AP function and allowed cytoplasmic disulfide bond formation. These mutations mapped to thioredoxin reductase, encoded by the

trxB gene. Efforts to improve disulfide bond formation in the cytoplasm created strains deficient in both trxB and glutathione oxidoreductase (gor) which did improve disulfide formation but resulted in a growth defect that rendered cells less than useful for protein expression applications. Further refinement led to the creation of a trxB gor ahpC* triple mutant known as FA113 that permits formation of stable disulfide bonds in the cytoplasm and grows comparably to wild-type [17]. Using strain FA113, successful cytoplasmic expression of oxidized, biologically active tissue plasminogen activator, a complex protein containing 17 disulfide bonds, was achieved [17]. A related strain, namely DR473, that carries a genomic copy of the trxB gene under control of the arabinose-inducible promoter pBAD, was developed to enable greater tuning of the redox environment of the cytoplasm. These redox engineered strains (see Table 1.1 for relevant genotypes/phenotypes) have been used extensively for the production of proteins with multiple disulfide bonds [53, 168]. The full, detailed story of the creation of strains with oxidizing cytoplasm can be found in a recent review/retrospective by Jon Beckwith [13].

Molecular chaperones

The role of molecular chaperones including DnaK, DnaJ, GrpE, GroEL and GroES as well as their mechanism of action was extensively reviewed by Houry in 2001 [98] and thus will not be covered in detail here. Briefly, chaperones in the cytoplasm can serve three purposes: (i) holding chaperones (e.g., IbpB) that retain partially folded proteins until the arrival of folding chaperones and help shield their substrates from proteolysis; (ii) folding chaperones (e.g., DnaK and GroEL) that rely on ATP to drive either folding in the case of partially folded

substrates or unfolding in the case of misfolded substrates; and (iii) unfolding chaperones (e.g., ClpA and ClpB) that help to solubilize proteins that have aggregated and/or unfold them in preparation for refolding (by a folding chaperone) or degradation (by a protease). Following their initial interaction with trigger factor (see above), nascent chains are released from the *E. coli* ribosome and can either attain folded conformations unassisted or might require the activity of additional molecular chaperone systems. At an early stage of folding, unfolded polypeptide segments are typically recognized by the Hsp70 chaperone system, which is comprised of DnaK, DnaJ and GrpE [30]. The chaperone system formed by GroEL and GroES typically becomes involved at a later stage in the folding process. This system is arguably one of the most significant chaperone complexes in *E. coli* as evidenced by the fact that it is the only such system that is essential under all conditions of growth [64] and by the extraordinary number of proteins that interact with GroEL/GroES in their lifetime [99, 115]. Many chaperones, including GroEL/GroES, exhibit broad substrate specificity for a diverse range of newly synthesized proteins [99, 115], including many heterologous proteins that may not normally encounter the chaperone(s) in nature. As a result, overexpression of molecular chaperones has proven to be a useful strategy for improving cytoplasmic folding and solubility of a wide of array of protein substrates that bear little resemblance to one another with respect to sequence, structure and function [86, 140, 172, 173, 226, 173, 227]. However, despite this success, chaperone overexpression strategies frequently require tedious trial-and-error analysis to optimize the choice and expression level of the chaperone system(s) for each new target protein under investigation. This inherent lack of specificity for a given protein effectively limits the folding efficiency of a chaperone towards a specific target [62]. To address this limitation,

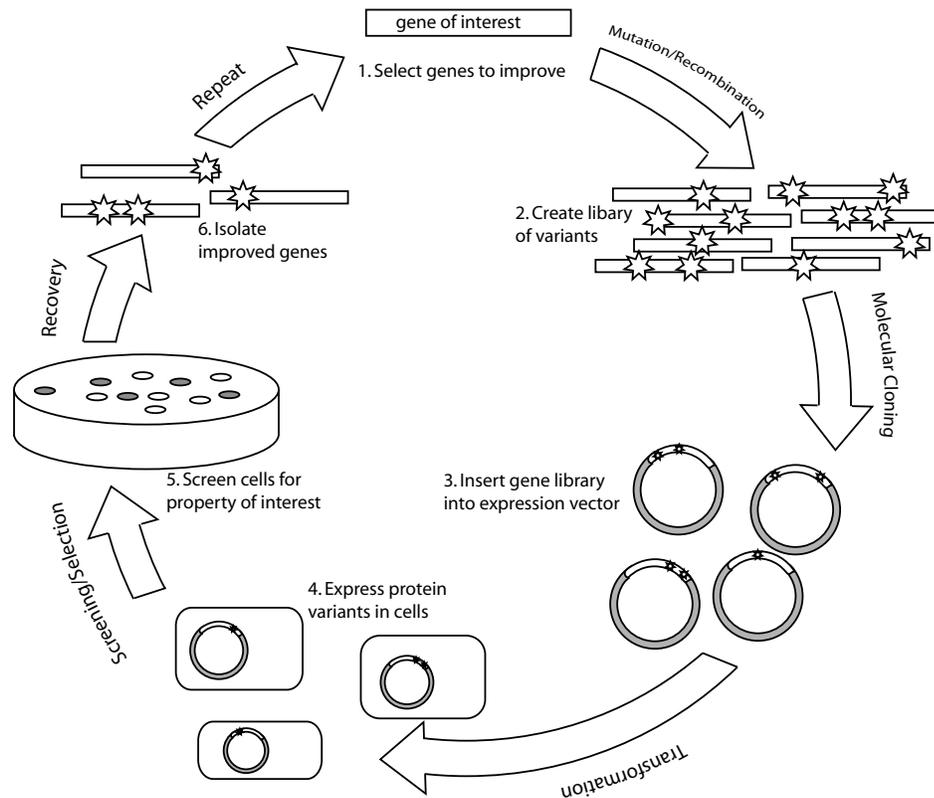


Figure 1.3: Schematic of the directed evolution process. By following an iterative cycle of DNA mutagenesis to introduce diversity and selection to identify unique clones with the desired phenotype, researchers can obtain proteins (or protein complexes) with improved or even vastly different functions.

Weissman and coworkers used a directed evolution strategy (see Fig. (3)) to create a variant of GroEL/S that dramatically enhanced folding of a single substrate [243]. This acquired substrate specificity came at the expense of the chaperone systems generality as manifested by the inability of the GroEL/S variants to fold a variety of their natural substrates. Nonetheless, this groundbreaking study was the first example in what will undoubtedly become a growing pursuit of researchers in the future: chaperone engineering.

Cytoplasmic protein folding reporters

The development of tailor-made chaperones specific for certain substrates, especially by directed evolution, necessarily requires the engineering of robust genetic screens or selections to identify factors that effect folding of the target substrate. This can be particularly challenging since relatively few substrates, especially those which are foreign to host, are screenable. Thus, a suitable assay for proper folding is essential for substrates such as these as well as for the identification of new engineer-able targets in the protein folding landscape. Fortunately, a variety of genetic tools have been developed to identify properly folded proteins in the bacterial cytoplasm. In general, these assays involve construction of a genetic fusion between an analyte protein-of-interest and a reporter protein. The underlying premise is that the folding behavior of the analyte protein will directly influence the reporter protein such that the reporter activity is representative of the analytes folding status. For example, Waldo and colleagues fused green fluorescent protein (GFP) to the C-terminus of a wide array of test proteins and found that bacterial cell fluorescence correlated positively with the relative solubility of the analyte partner [240]. A number of similar approaches have been developed using either chloramphenicol acetyltransferase [156], the Gal11P-activation domain [6], the RNase-A S-peptide [113] or dihydrofolate reductase [146] as the reporter protein in the fusion construct; sometimes even in combination [104]. More recently, next-generation protein folding assays have been developed that take advantage of the structural complementation of split protein fragments derived from GFP [34, 33], β -galactosidase [250], and dihydrofolate reductase (DHFR) [180] instead of full-length reporter proteins. The advantage of these reporters, split or otherwise, is that each is amenable to high-throughput experimentation thereby enabling combinatorial strategies for

evolving better folding proteins or cells [152, 151, 241]. It should be noted that an implicit assumption in assays involving protein fusions is that the analyte protein must affect the folding and function of the reporter protein, but not vice versa. Unfortunately, this is not always the case as reporter proteins can often attain an active conformation even when the analyte to which they are fused is insoluble [184] or aggregated [230]. Thus, care must be taken when interpreting results from experiments where these approaches have been employed to characterize or engineer the protein folding machinery in cells. An alternative method for assaying protein folding was recently developed that does not rely on the expression of artificial protein chimeras, but rather on the genetic response of the cell to misfolded proteins [130]. Cells have evolved exquisite methods for exerting protein folding quality control including, for example, the upregulation of many chaperones and proteases in response to misfolded or aggregated protein. Thus, simple cloning of a reporter gene (e.g., *lacZ*) under the control of a promoter for such an upregulated chaperone or protease results in a biosensor that responds to protein misfolding. To date, genetic elements that have been used to successfully assay protein folding in this manner include the promoter region of the small heat shock protein *ibpAB* fused to *lacZ* [141] and the promoters from *ibpAB* and *fxsA* linked in tandem (*ibpfxs*) upstream of the luciferase reporter gene, *lucA* [130]. While this promoter trap system should clearly be useful for evolving protein sequences directly, it may be quite challenging to explore the effect of cellular factors, especially heat shock chaperones, that may themselves activate the promoter. In addition, it is not clear if such a system will be responsive to a broad spectrum of different protein structures and misfolding events. A final drawback is the fact that productive folding results in the absence of any signal, which may prove problematic in

library screening experiments where improved folding is the desired outcome. Recently, we developed a protein folding selection that relies on both a fusion to a reporter protein and the use of cellular folding quality control [71]. The basis for this assay is the observation that protein transport through the bacterial twin-arginine translocation (Tat) pathway depends on correct folding and solubility of the substrate protein prior to transport [53]. In this system, a test protein is expressed as a sandwich fusion between an N-terminal Tat signal peptide and a C-terminal TEM1 β -lactamase (Bla) reporter protein. Bla confers antibiotic resistance and is only effective when present in the periplasm. Viability of *E. coli* cells on antibiotic selective media expressing tripartite chimeras was shown to dependent on the solubility of the test protein. Thus, by exploiting a novel cellular folding quality control mechanism as well as the activity of a reporter enzyme that is amenable to high-throughput selections, we expect this reporter assay will be quite useful for identifying and engineering cytoplasmic factors that modulate the protein folding landscape.

Engineering membrane transport machinery

Not all cellular proteins, however, are destined to remain at their site of synthesis in the cytoplasm; approximately 20% of the proteome is secreted across the cytoplasmic membrane [188]. In fact, one way in which bacteria handle proteins that are incapable of folding in the cytoplasm is by simple export to allow folding in a different cellular compartment, such as the inner membrane or the periplasm. Proteins bound for transport usually contain an N-terminal signal peptide that promotes interactions with general and pathway-specific chaperones and also directs the protein to interact with the translocation machinery

(for a review, see [70]). Most proteins in *E. coli* are transported via the general secretory (Sec) pathway [188], although several other functionally distinct pathways for inner membrane transit have been identified that provide intriguing targets for exploiting the cellular protein folding landscape.

Sec-dependent transport

The bulk of protein transport across the inner membrane of bacteria (and also the cell wall of Gram-positive bacteria) occurs via the Sec secretion pathway [188]. This pathway identifies secretion substrates by cleavable N-terminal peptide extensions, known as signal peptides, which target a protein substrate for transport. Proteins traverse the membrane through the Sec translocon, a complex of three integral membrane proteins (SecY, SecE and SecG) which form a funnel structure that constricts to a ring of 5-8 nm in diameter [235]. As a result of this narrow conduit, proteins transported through the SecYEG translocon remain largely unfolded, possibly resembling a random coil, and effectively ratchet their way through the pore in a process requiring SecA-mediated ATP hydrolysis [198, 217]. In contrast, proteins that have adopted a kinetically stable, folded conformation are precluded from transport via the Sec pathway [189]. For a subset of native Sec-targeted proteins, the dedicated molecular chaperone SecB helps to maintain the substrate in an extended, export-competent conformation. It is particularly noteworthy that protein secretion for biotechnology purposes has relied almost exclusively on the Sec pathway. The significance of the Sec pathway is highlighted by the very extensive literature covering various leader peptides useful for recombinant protein expression as well as by the enterprise encompassed by technologies for combinatorial library screening

of proteins and peptides such as yeast surface display [19] (Invitrogen Corporation) and phage display [210] (Dyax Corporation). However, the fact that polypeptides secreted by Sec must be largely unfolded prior to transport has hindered its use for certain applications. Fortuitously, several groups have made use of prl suppressor mutants (that map in several sec genes [61]) for improved Sec-dependent expression of recombinant proteins [154, 181, 193], even though these mutants were not originally isolated for such biotechnological purposes. Prl suppressors are known to allow significant export of substrates with an array of defective, or even deleted, signal peptides [54, 75, 187]. This lack of allele specificity has led to the suggestion that the prl mutations broaden the selectivity of the translocase [61] and, as a result, promote increased export efficiency of Sec-targeted heterologous proteins [154, 181, 193]. Another effective strategy for improving Sec transport efficiency is the overexpression of SecB, which can markedly increase the quantity of human recombinant proteins localized in the *E. coli* periplasmic space [15].

SRP-dependent transport

In addition to SecB-dependent transport, *E. coli* cells also possess an SRP-dependent pathway for targeting to the SecYEG translocon [150, 234]. Recognition of the N-terminal signal peptide by the SRP induces a change in the ribosome that promotes interaction with the SRP-receptor (SR) FtsY and targets the SRP/SR/ribosome complex to SecYEG. A key aspect of the SRP pathway is that substrate proteins are transported into or across the inner membrane simultaneously with translation, thereby ensuring that secondary structures are not formed in the cytoplasm and eliminating the inhibitory effect of protein fold-

ing on export. Thus, in addition to translocating plasma membrane proteins such as the lactose permease [150], the bacterial SRP pathway is also responsible for export of certain soluble periplasmic proteins that fold rapidly in the cytoplasm and thus require co-translational export [100]. It should be noted that while SRP-dependent transport utilizes the SecYEG translocon, some evidence suggests that it may also use the SecA ATPase to propel the nascent polypeptide across the membrane [199]. Based on these observations, the SRP pathway should eventually prove to be a useful mechanism for efficient heterologous expression of notoriously difficult membrane proteins or very stable and fast-folding non-membrane proteins. Along these lines, a recent study by Plckthun and coworkers demonstrated that the display efficiency of a fast-folding protein on filamentous phage could be improved up to 700-fold simply by replacing Sec- with SRP-dependent signal peptides [215]. For expression of membrane proteins, it may prove useful to combine SRP targeting with an *E. coli* host strain such as C41(DE3) or C43(DE3) (see Table 1.1) that has been engineered for high-level expression of certain membrane proteins [163].

Tat-dependent transport

The bacterial Tat pathway is a Sec-independent mode of inner membrane transport that is perhaps best known for its remarkable ability to transport proteins that have undergone folding prior to translocation (for a recent review of the Tat system, see [139]). While many of the mechanistic details remain a mystery, our recent studies have revealed the existence of folding quality control mechanism that governs the transport of synthetic Tat substrate proteins of both prokaryotic and eukaryotic origin [53, 71, 118]. As folding quality control is a crucial

aspect of protein expression and engineering, the Tat system should emerge as an ideal platform for numerous biotechnology applications. Already it has been used in *E. coli* for expressing recombinant therapeutics such as antibodies and human tissue plasminogen activator [53, 118, 192], screening and engineering both protein solubility [71] and protein-protein interactions [216, 244] and displaying proteins on phage [177, 224]. A current drawback of the Tat system is its relatively low transport efficiency, however, identification of optimized signal peptides [52, 231] as well as factors (e.g., phage shock protein A, DnaK/DnaJ) whose overexpression increases translocation efficiency [51, 182] should help to alleviate this bottleneck.

Engineering protein folding in the periplasmic space

General properties of the periplasm. The periplasmic environment differs in many ways from that of the cytoplasm. First, the protein concentration is much lower than that of the cytoplasm, so molecular crowding effects are likely not as pronounced. However, macromolecular crowding has been shown to affect the folding properties of certain proteins in this compartment compared to in dilute solution [160]. Second, the environment of the periplasm is much more susceptible to the conditions in the extracellular environment such as pH, temperature, and concentration gradients of small molecules [165]. For these reasons, purification of proteins from the periplasm is less complex than purification from whole cell extracts, and methods whereby proteins are extracted from the periplasm without whole cell lysis have been used extensively in the literature for many years [171]. Third, the lack of a significant ATP concentration in the periplasm precludes classical chaperone activity (which entails ATP-

driven folding). Fourth, the oxidation-reduction potential of the periplasm is such that disulfide bonds can form readily in proteins containing multiple cysteine residues. Such bonds generally increase the stability of proteins, thus the formation of disulfides is a critical step along the folding pathway of many native and non-native proteins expressed in the *E. coli* periplasm.

Engineering the redox environment of the periplasm.

Periplasmic disulfide bond formation is catalyzed by a family of protein disulfide isomerases (PDIs), all of which belong to the thioredoxin superfamily. The most notable of these, DsbA, is oxidized by its membrane-bound partner, DsbB, and uses this oxidation potential to catalyze the formation of disulfide bonds (the reader is referred to [107, 165] for a more detailed review of the mechanism of disulfide bond formation in the periplasm). Given that disulfide bond formation can greatly affect folding and stability of proteins and also that eukaryotic proteins with therapeutic potential frequently have a high cysteine content, it is not surprising that strains of *E. coli* have been engineered that allow exquisite control of disulfide bond formation in the periplasm. For instance, overexpression of PDI family members such as *E. coli* DsbA and DsbC has proven to be a reliable means of improving the folding yield of proteins with multiple disulfide bonds [106, 136]. Alternatively, introduction of a *dsbA::kan* allele into strain DR473 (resulting in strain DRA, see Table 1.1) resulted in a strain that exhibited an inverted redox pattern (i.e., oxidizing cytoplasm, reducing periplasm) when glucose was added to repress *trxB* expression and a *dsbA* phenotype (reducing cytoplasm, reducing periplasm) when arabinose was added to induce *trxB* [53]. Thus, the combination of strains DR473 and DRA provides researchers with a

Table 1.1: Engineered *E. coli* strains with altered folding environments

Strain Name	Genotype	Phenotype	Reference or source
DHB4	MC1000 <i>phoR</i> $\Delta(\text{phoA})$ PvuII $\Delta(\text{malF})_3$ F' <i>lacIq</i> ZYA <i>pro</i>	Reducing cytoplasm	Beckwith lab
FA113	DHB4 <i>trxB::kan</i> <i>gor552</i> Tn10Tet <i>ahpC*</i>	Oxidizing cytoplasm	[17]
DR473	DHB4 ΔtrxB <i>gor552</i> Tn10Tet <i>ahpC*</i> Tn10Cm (<i>araC</i> Para- <i>trxB</i>)	Arabinose-inducible oxidizing cytoplasm	Beckwith lab
DHA	DHB4 <i>dsbA::kan</i>	Reducing periplasm	[53]
DRA	DR473 <i>dsbA::kan</i>	Arabinose-inducible oxidizing cytoplasm and reducing periplasm	[53]
BL21	B F- <i>ompT</i> <i>hsdSβ</i> (<i>rβ-mβ-</i>) <i>gal</i> <i>dcm</i>	<i>OmpT</i> and Lon protease-deficient	Novagen
BL21(DE3)	BL21 carrying the DE3 lambda lysogen	As BL21 with T7 polymerase for pET vector expression	Novagen
Rosetta	BL21 + pRARE	As BL21 with plasmid supplying tRNAs for seven rare codons (AUA, AGG, AGA, CUA, CCC, CGG, and GGA)	Novagen
C41(DE3), C43(DE3)	BL21(DE3) supp	As BL21(DE3) with unknown suppressor mutations that permit growth of cells following induction of certain toxic (e.g., membrane) proteins	[163]
HM130	KS272 (wild type) <i>degP</i> <i>ptr</i> <i>ompT</i> <i>tsp</i> <i>eda</i>	Deficient in all cell envelope proteases	[161]

powerful genetic system for toggling the redox state of both the cytoplasm and the periplasm To observe and optimize folding of proteins in both of the bacterial compartments.

Chaperone activity in the periplasm.

Since the periplasm is not enriched in ATP like the cytoplasm [11], the existence of ATP-mediated chaperone systems in the periplasm is unlikely. Using a screen based on the genetic response to protein misfolding in the periplasm, Raina and coworkers identified three proteins (SurA, FkpA and Skp) that function as periplasmic folding catalysts [106]. Shortly thereafter, Bothmann and Plckthun employed a clever strategy based on phage display of a single-chain antibody fragment (scFv) that tended to aggregate in the periplasm to screen an *E. coli* genomic library for factors that could enhance scFv folding in the periplasm and independently identified Skp [22]. The FkpA and SurA proteins are peptido-prolyl cis/trans isomerases (PPIases) that contribute to folding of unfolded intermediates by stabilizing the isomerization of prolines in the polypeptide chain. In addition, FkpA has a chaperone activity independent of its PPIase activity [4]. Skp has been characterized as a molecular chaperone that interacts with unfolded proteins as they emerge in the periplasm from the Sec translocation machinery [106] and appears to make the broadest contribution to protein folding in the periplasmic space. As with many of the molecular chaperones discussed so far, overexpression of these three proteins has led to greater expression and in some cases, reduced aggregation and formation of inclusion bodies in the periplasm [22, 23, 92]. In addition to periplasmic chaperone overexpression, Georgiou and coworkers have recently explored the notion of chaperone engineering in the periplasm. One such approach explored directed evolution of DsbC to improve its foldase and/or isomerization activity towards a complex, multidisulfide substrate protein that was poorly expressed in the periplasm [137]. A key aspect of this study was the development of a phage display assay linking the genotype of the foldase (encoded on a phagemid) to

the phenotype of the target substrate (displayed as a fusion to pIII on the head of the phage). Improved substrate folding was then assayed by the ability of the substrate to bind to a surface-immobilized ligand. Thus, an obvious limitation of this method is that it can only be used to screen for chaperones whose substrate proteins exhibit activities (i.e., ligand binding or fluorescence) that are amenable to high-throughput assays. A second approach employed a rational design strategy to engineer DsbA-DsbC chimeras that could restore protein oxidation in a *dsbA* background and also fully substitute for DsbC in disulfide-bond rearrangement [204]. Remarkably, several of the engineered chimeras could fulfill both of these functions, acting as a single catalyst for both disulfide-bond formation and rearrangement.

Periplasmic protein folding reporters.

Since periplasmic expression is favorable for many proteins, especially those requiring the formation of disulfide bonds (including nearly all antibodies and fragments derived thereof), a useful addition to the protein engineers toolkit would be a reliable reporter of protein folding in the periplasm. A system to report periplasmic folding could conceivably identify new factors that affect protein folding in trans (e.g., chaperones) in a manner that does not depend on the activity of the target protein [22]. To our knowledge, however, there are currently no general folding reporters of this type reported in the literature. One possible factor contributing to this dearth of periplasmic folding reporters is the apparently unreliable nature of certain reporter enzymes. For instance, Ari and coworkers created fusions between several reporter enzymes and soluble and insoluble variants of the *E. coli* maltose binding protein to study the relation-

ship between periplasmic protein folding and aggregation [3]. These authors suspected that reporter enzymes fused C-terminally to the aggregation-prone MBP variant, known as MalE31, would be incorporated into inclusion bodies and therefore be inactive. Upon testing these fusions, they observed that enzymes fused to MalE31 accumulated in the insoluble fraction but unexpectedly retained their catalytic activities. Based on this observation, they concluded that reporter enzymes were not a reliable indicator of protein solubility in the periplasm. Another possible factor contributing to the lack of periplasmic folding reporters is the inability of GFP, one of the most commonly used reporter proteins, to attain an active fluorescent conformation when transported to the periplasm via the Sec machinery [65]. Interestingly, when exported through the Sec pathway, GFP is detected in the soluble fraction of cell lysate, yet fails to form the fluorescent chromophore for reasons that are unclear. Recently, Sorenson *et al.* reported that the use of co-translational translocation allowed an engineered variant known as superfolder GFP to attain fluorescence in the periplasmic space [212]. Superfolder GFP has been used in the past as a folding enhancer to proteins, but given its highly structured and stable nature, it is still unclear whether sfGFP could function as a reporter of protein folding. In any case, the development of a protein folding reporter for the periplasm would contribute greatly to our understanding of the complexities of the periplasmic folding environment, and could be instrumental in isolating additional yet-to-be identified factors that contribute to better periplasmic expression of recombinant proteins.

Engineering proteolysis

The last stage in the life of a protein is proteolytic degradation into constituent amino acids (for a complete review of proteolysis in *E. coli* see [85]). Proteases are ubiquitous throughout the cell and are essential to cellular quality control. However, they can be troublesome during overexpression of recombinant proteins, especially if the target protein is prone to aberrant folding. To tackle this issue, Meerman and Georgiou constructed a family of 25 isogenic strains deficient in all known cell envelope proteases including DegP, Protease III, Tsp(Prc), and OmpT to select suitable hosts based on the optimum reduction in protease activity [161]. The genotype of HM130, one of the most effective strains relative to its growth rate, is listed in Table 1.1. Similarly, one of the most popular strains for recombinant protein expression is BL21, a derivative of the *E. coli* B strain that is deficient in the cytoplasmic protease Lon and the periplasmic OmpT protease (see Table 1.1). In contrast to its negative effect on protein expression, intracellular proteolysis can also be productively harnessed for a number of different purposes. For instance, Sauer and colleagues have designed a series of modified SsrA degradation tags that have varying affinities for the protease ClpXP [159]. Insertion of these tags into cellular proteins enables synthetic toggling of the tagged proteins degradation rate. In another example, the same SsrA degradation tag was used to ensure exclusive localization of Tat-targeted GFP solely to the periplasm as non-transported GFP was efficiently degraded in the cytoplasm by ClpXP [52]. Finally, proteases have also been the focus of directed evolution studies in which the reactivity of the protease is tailored for novel substrates, while simultaneously excluding reactivity toward undesired substrates. Specifically, screening of a large library of random mutants of the *E. coli* endopeptidase OmpT resulted in the isolation of a variant that cleaved an

Ala-Arg peptide bond instead of the Arg-Arg bond initially preferred by the enzyme with a greater than 3×10^6 -fold selectivity of Ala-Arg over Arg-Arg [236].

Introduction to bacterial *N*-linked glycosylation

Protein glycosylation is a post-translational modification made to proteins wherein sugars (e.g. glucose, mannose, galactose, etc.) are covalently linked to proteins. In eukaryotes, this process occurs in the endoplasmic reticulum (ER) and Golgi bodies before the mature protein is released into the cytoplasm [94]. Two types of protein glycosylation exist: *N*-linked and *O*-linked. *O*-linked glycosylation, in which carbohydrate groups are transferred to serine or threonine residues in the Golgi, is GCSF folding. *N*-linked glycosylation, however, has been shown to be very important to many protein functions and properties [251].

N-linked protein glycosylation involves the attachment of oligosaccharide groups to asparagine residues of nascent polypeptides. In yeast and higher eukaryotes, a conserved sequence of Asn-X-Ser/Thr, where X is any amino acid but proline, is the consensus sequence for *N*-linked glycosylation. Glycosylation first occurs in the ER, where the oligosaccharide is attached to the nascent polypeptide before it adopts a folded conformation [93]. A core oligosaccharide common to all eukaryotes is attached in the ER and removal and addition of different monosaccharides in the Golgi varies from organism to organism. For example, the most common yeast glycoform is mannose-rich, while the human glycoform incorporates sugars like galactose and N-acetylneuraminic acid [251]. Glycosylation has been shown in many cases to enhance folding

of glycoproteins [103]. In fact, it has been speculated that the addition of N-glycans has been an essential factor that has allowed eukaryotes to express high amounts of complex proteins that are well-folded [93]. Many proteins that are naturally glycosylated will express poorly or not at all in the absence of glycosylation [251]. Until the mid-1990s, it was widely believed that protein glycosylation was only present in eukaryotes [14]. Evidence of N-linked protein glycosylation in the genome of the pathogenic bacterium *Campylobacter jejuni* was found in 1999 [222] and the presence of glycosylated proteins in *Campylobacter* was demonstrated by Linton and colleagues in 2002 [144]. Not long after, the so-called *pgl* (protein glycosylation) genetic locus of *C. jejuni* was successfully transferred into *E. coli*, and it was shown that *E. coli* cells were capable of glycosylating proteins using the glycosylation machinery of *C. jejuni* [239]. In addition, the structure of the heptasaccharide was revealed by NMR spectroscopy to be GalNAc- α 1,4-GalNAc- α 1,4-[Glc β 1,3]GalNAc- α 1,4-GalNAc- α 1,4-GalNAc- α 1,3-Bac- β 1,N-Asn, where Bac is bacillosamine, 2,4-diacetamido-2,4,6-trideoxyglucose attached by the engineered *E. coli* was found to be identical to that of *C. jejuni* [145]. In 2006, each gene in the 17-kb *pgl* locus was individually characterized by Kelly and colleagues [114]. The locus consists of 13 genes, shown in **Figure 4.1**, most of the products of which participate in glycosylation. The process proceeds via 5 general steps: (1) saccharide and sugar-nucleotide biosynthesis in the cytoplasm, (2) attachment of the base sugar (bacillosamine) to lipid precursors on the membrane, (3) transfer of sugars to the glycan on the membrane, (4) "flipping" of the glycan from the cytoplasmic surface to the periplasmic surface of the inner membrane, and (5) transfer of the oligosaccharide to the asparagine residue at the glycosylation site of the protein.

The laboratory of Markus Aebi (Institute of Microbiology, ETH Zurich,

Switzerland) has pioneered the functional study of the *C. jejuni* machinery in *E. coli*. Discoveries from this group include the requirement of the D/E-X₁-N-X₂-S/T, (where X₁ and X₂, are any amino acid except proline) glycosylation consensus sequence [129, 174], which is more specific than the classical eukaryotic acceptor sequence of N-X-S/T. Additionally, protein folding studies on bovine ribonuclease A in this group have concluded that the oligosyltransferase PglB can glycosylate proteins after they have attained a folded structure, but glycosylation is much more efficient when the substrate protein remains unfolded [128]. Also, it was discovered that PglB has relaxed substrate specificity to glycans, as it was shown to successfully attach the O antigen of *E. coli* to protein[67]. This relaxed specificity for glycan was supported by the in vitro experiments of Glover and colleagues with various undecaprenyl-linked disaccharides [82]. *While PglB has been shown to have relaxed specificity, it is unknown if it is capable of attaching eukaryotic- and human-like glycans to proteins.*

Conclusions and future directions

From translation of a nascent polypeptide to its eventual folding, researchers have made considerable advances in the study and engineering of cellular environments with respect to protein folding. Our analysis of efforts to modify the protein folding landscape has shown that strategies for better folding of recombinant proteins are being pursued in cytoplasmic and extracytoplasmic compartments of the cell, including the inner membrane, and periplasm. Despite the production of disulfide-bonded proteins in the cytoplasm, the periplasm is still the optimal location for folding and expression of complex proteins. This notion is only strengthened by the recent demonstrations that fully functional

immunoglobulin G (IgG) molecules can be expressed [208] and engineered [157] in the periplasm of *E. coli*. In addition, the discovery of a bacterial system of *N*-linked protein glycosylation in the pathogenic bacterium *Campylobacter jejuni* and its functional transfer into *E. coli* [239] is especially encouraging. In this system, the bacterial oligosaccharide is produced in the cytoplasm, but transfer of the sugar group onto protein occurs in the periplasm. Thus, with continued advances in the engineering of translation, folding, secretion and post-translational modification machinery in *E. coli*, it may not be long before biosynthesis of (nearly) authentic human glycoproteins in this simple host is routine practice.

CHAPTER 2

A RAPID PROTEIN FOLDING ASSAY THE BACTERIAL PERIPLASM

Introduction

¹Since the inception of recombinant DNA technology, maximizing the solubility of heterologous proteins in *Escherichia coli* has been a principal goal of modern biotechnology[11, 219, 80]. Despite the numerous advances in this area over the past three decades, high-level expression of correctly folded, soluble proteins for laboratory and preparative purposes remains a significant challenge. Indeed, following their expression in the cytoplasm of *E. coli*, many recombinant proteins of prokaryotic and eukaryotic origin are prone to misfolding[56, 148, 26] and are subsequently degraded by cellular proteases[9, 228] or deposited into inactive inclusion bodies[24, 238]. Protein misfolding in the cytoplasm can be a consequence of the relative crowdedness of this compartment where macromolecule concentration can reach 300-400 mg/ml[58]. In addition, post-translational processing steps such as disulfide bond formation[107] or *N*-linked glycosylation,[248] which are often required for correct folding, are absent from the cytoplasm of wildtype *E. coli* cells. The periplasm in *E. coli*, on the other hand, is often the preferred compartment for protein expression because it: (1) contains significantly fewer proteins, particularly proteases, compared to the cytoplasm; (2) houses a network of redox enzymes that catalyze the formation and isomerization of disulfide bonds; and (3) permits *N*-linked glycosylation in *Campylobacter jejuni*[248] and *E. coli*[239]. Consequently, periplasmic proteins are often easier to isolate, less prone to crowding-induced aggregation and/or

¹Adapted with permission from Mansell TJ, Linderman SW, Fisher AC, and DeLisa MP, "A rapid protein folding assay the bacterial periplasm" (2010) *Protein Science*, **19** 5, pp. 1079-1090

proteolytic degradation and more efficiently folded compared to their cytoplasmic counterparts. However, even though the significance of the periplasm for protein expression is firmly established[165, 80], there are currently no genetic reporter assays for monitoring protein folding in this subcellular compartment. In stark contrast, numerous folding reporter systems have been developed for detecting correctly folded, soluble proteins in the cytoplasm of living *E. coli* cells[102, 156, 240, 250, 34, 71, 40]. These approaches commonly rely on a genetic fusion between a protein-of-interest (POI) and a reporter protein whose specific phenotype is independent of the POI's function[241]. In this scenario, when the POI folds into a soluble conformation, the reporter to which it is fused is functional. In contrast, when the POI misfolds or aggregates, the fused reporter is inactive and a null phenotype is observed. Another typical feature of these approaches is that they can be employed even when structural or functional information about the target is lacking. Perhaps the most useful aspect of these assays is that they can be combined with well-established methods for creating protein diversity libraries to screen or select for soluble variants of recalcitrant proteins[73, 71, 179, 240, 40, 33] with only one exception reported so far[184]. Hence, a screen or selection for monitoring protein folding in the bacterial periplasm would be a desirable accomplishment, but to date has been met with technical difficulties[3]. In this study, we developed an activity-independent selection strategy that reliably reports the "folding robustness" of POIs expressed in the periplasm. The term folding robustness is used here to denote both the chemical solubility related to correct folding of the POI and the avoidance of aggregation or degradation. The assay is based on a tripartite fusion between: (1) an N-terminal signal peptide from *E. coli* DsbA (ssDsbA), which has previously been shown to direct proteins through the bacterial signal recognition particle

(SRP)-dependent translocation pathway[199], (2) the POI, and (3) a C-terminal fusion of mature TEM-1 β -lactamase (Bla). We chose the SRP pathway because of its unique ability to perform co-translational export[150, 234] whereby nascent polypeptides bearing N-terminal SRP signals are directly translated across the inner membrane by a ribosome-translocon complex[185]. As a result, SRP-exported proteins are expected to experience little or no residence time in the cytoplasm, thus ensuring that protein folding takes place predominantly in the periplasmic space. Indeed, we observed that the antibiotic resistance of cells expressing engineered ssDsbA-POI-Bla chimeras correlated with the periplasmic folding behavior of the POI. Thus, simple selection on β -lactam antibiotics such as ampicillin (Amp) enabled discrimination between folded and misfolded conformations of POIs, even when (i) structural and functional information and (ii) activity assays for the POIs were lacking. The method was also capable of evaluating trans-acting factors that influence protein folding in the periplasm such as disulfide bond formation enzymes and molecular chaperones. Finally, the potential of the assay for improving periplasmic protein folding and solubility was demonstrated using a directed evolution strategy. Taken together, these results reveal our periplasmic protein-folding assay to be a powerful new experimental tool for elucidating the factors that affect the folding of proteins in this important biological compartment.

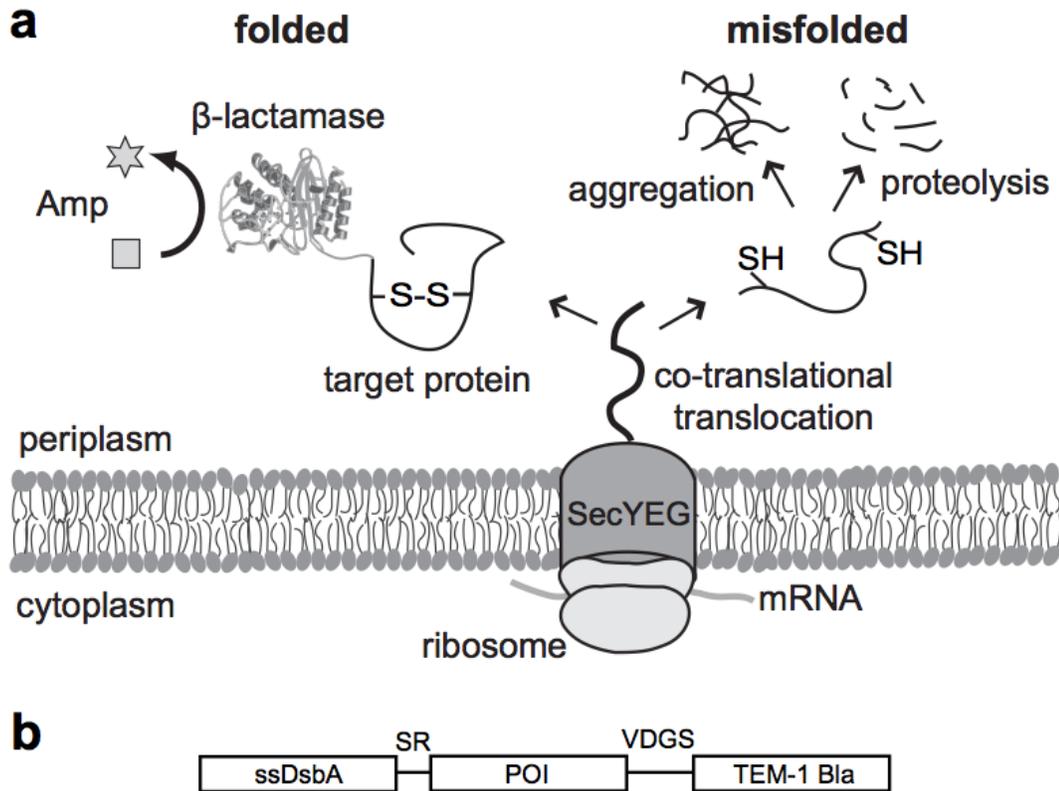


Figure 2.1: Schematic of SRP-mediated periplasmic folding reporter. (a) The DsbA signal peptide enables co-translational translocation across the cytoplasmic membrane via the SRP-dependent pathway. Nascent polypeptides enter the periplasm where they fold properly, conferring Amp resistance to cells, or misfold, leading to eventual aggregation and/or proteolysis. (b) Fusion proteins created in this study contained an N-terminal DsbA signal peptide (ssDsbA) followed by a POI and a C-terminal selectable marker (TEM-1 Bla). Short linker sequences (SR, VDGS) were created at the junctions.

Results

An engineered assay for folding and solubility in the periplasm.

To create a protein folding assay for the periplasm, we constructed a plasmid called pDMB that permitted expression of any POI as a sandwich fusion between an N-terminal SRP-dependent export signal (ssDsbA) and a C-

terminal reporter enzyme (Bla). Consistent with the logic underlying fusion-based folding reporters for the cytoplasm[156, 240, 250, 34], we reasoned that productive folding of the downstream Bla protein domain and consequent resistance to β -lactam antibiotics is directly related to the folding competence (avoidance of aggregation and inclusion body formation) of the upstream POI. Thus, correctly folded, soluble proteins would be expected to confer Amp resistance to *E. coli* cells and provide a selectable phenotype for periplasmic protein folding (Fig. 2.1). We chose the ssDsbA signal peptide because it is known to direct co-translational export of heterologous proteins through the SRP pathway[199, 215] and thus should effectively partition ssDsbA-POI-Bla fusions to the periplasm with little to no residence time in the cytoplasm. In this study, 32 different POIs were cloned into pDMB and characterized based on their resistance to Amp as described below.

Monitoring DsbA-dependent folding of enzymes and antibodies in the periplasm.

As proof-of-concept, we first investigated whether our assay could report the folding behavior of *E. coli* alkaline phosphatase (PhoA). We chose PhoA because, like many other extracytoplasmic proteins, its correctly folded, catalytically active conformation is dependent upon two disulfide bonds that are formed by the primary periplasmic oxidant DsbA[211]. Expression of ssDsbA-PhoA-Bla in wild-type DHB4 cells conferred resistance to 100 mg/ml Amp as evidenced by spot plating (Fig. 2.2a). When the same construct was expressed in DHA cells, which lack DsbA, resistance to this concentration of Amp was no longer observed. To quantify this difference in Amp resistance, single colonies

of DHB4 and DHA were challenged on increasing levels of Amp and values for the minimum bacteriocidal concentration (MBC) were determined. The MBC for DHB4 cells expressing ssDsbA-PhoA-Bla was 33-fold higher than that measured for DHA cells expressing the same fusion (Table 2). Plating on LB agar lacking Amp revealed no measurable growth difference between the DHB4 and DHA cells (Fig. 2.2a), confirming that the difference in Amp resistance was due to the DsbA-dependent oxidation of protein thiols required for PhoA folding. To confirm that this difference was not attributable to redox-dependent changes in the folding of the Bla moiety itself, cells expressing ssDsbA fused directly to Bla were similarly plated on Amp. Consistent with earlier findings[77], the periplasmic redox state had no measurable effect on the catalytic activity of Bla as evidenced by the identical resistance phenotypes for DHB4 and DHA cells expressing ssDsbA-Bla (Fig. 2.2a and Table 2). For comparison, we evaluated cytoplasmic folding of PhoA in DHB4 cells using a previously developed assay based on the twin-arginine translocation (Tat) protein export pathway[71]. In this system, POI-Bla fusions are targeted by an N-terminal ssTorA signal peptide to the Tat pathway, which requires that protein substrates fold correctly in the cytoplasm prior to transport across the inner membrane[53]. Thus, the efficiency of ssTorA-POI-Bla export, and the corresponding Amp resistance phenotype, is regulated by the folding behavior of the POI in the cytoplasm[73, 71]. As expected, we observed that DHB4 cells expressing ssTorA-PhoA-Bla were sensitive to 100 $\mu\text{g}/\text{mL}$ Amp (Fig. 2.2a). This Amp sensitivity was due to the fact that the cytoplasm of DHB4 cells is a reducing environment that renders PhoA misfolded and incompetent for export transport via the Tat pathway. Western blot analysis of the cytoplasmic and periplasmic fractions derived from DHB4 and DHA cells expressing the ssDsbA and ssTorA fusions confirmed that sol-

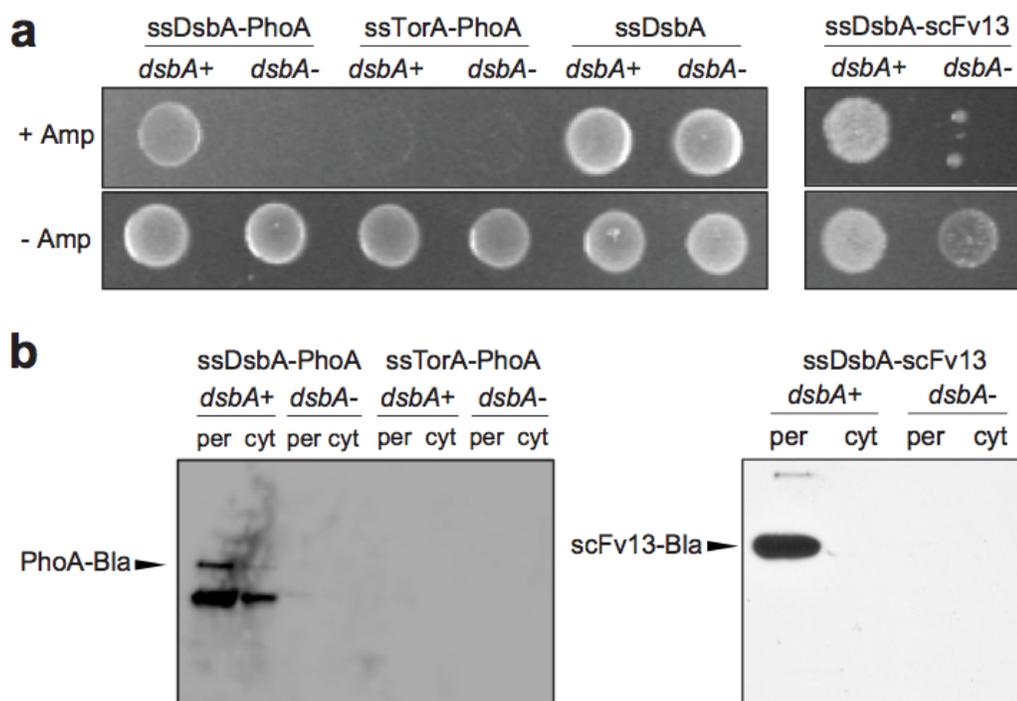


Figure 2.2: Redox-dependent folding reported via SRP selection. (a) An equivalent number of DHB4 (*dsbA+*) and DHA (*dsbA-*) cells expressing either ssDsbA-PhoA-Bla, ssTorA-PhoA-Bla, ssDsbA-Bla or ssDsbA-scFv13-Bla were spot-plated on LB/agar containing either 50 $\mu\text{g/ml}$ Cm (- Amp) or 100 $\mu\text{g/ml}$ Amp (+ Amp). (b) Western blot analysis of periplasmic (per) and cytoplasmic (cyt) fractions derived from DHB4 (*dsbA+*) and DHA (*dsbA-*) cells expressing either ssDsbA-PhoA-Bla, ssTorA-PhoA-Bla or ssDsbA-scFv13-Bla. Samples were blotted with Bla-specific antibodies.

uble expression of PhoA-Bla corresponded directly with the respective Amp phenotypes (Fig. 2.2b). This analysis also revealed that ssDsbA-PhoA-Bla was localized exclusively in the periplasm, consistent with a co-translational mode of export. Proteolytic degradation of the fusion was observed in the periplasm and to a lesser extent in the cytoplasm, but as discussed below, it does appear to affect the assay readout for PhoA and does not occur for most of the other proteins tested.

To determine whether the ability to report DsbA-dependent folding was a

Table 2.1: Amp resistance conferred by different target proteins

Protein of interest	Strain	MBC ($\mu\text{g/ml}$)	MIC ($\mu\text{g/ml}$)
alkaline phosphatase			
PhoA	DHB4	100	50
PhoA	DHA	3	3
ssTorA-PhoA-Bla	DHB4	6	6
ssTorA-PhoA-Bla	DHA	12	6
Bla	DHB4	3200	1600
Bla	DHA	3200	1600
anti- β -gal scFv13			
scFv13 (wild-type)	DHB4	400	200
scFv13 (wild-type)	DHA	50	50
scFv13-R4	DHB4	400	400
scFv13-R4	DHA	100	50
maltose-binding proteins			
MalE (wild-type)	DH5 α	250	200
G32D	DH5 α	200	175
I33P	DH5 α	200	175
MalE31	DH5 α	175	175
GroEL substrates			
MetF	DH5 α	25	25
MetK	DH5 α	6	3
amyloid- β peptides			
A β 42 (wild-type)	DH5 α	12	12
A2	DH5 α	25	25
A4	DH5 α	25	25
B12	DH5 α	25	25
B9	DH5 α	25	25
H2	DH5 α	50	25
GM6	DH5 α	100	100
GM7	DH5 α	100	50
GM11	DH5 α	100	50

general feature of our assay, we next analyzed the behavior of a single-chain antibody fragment specific for β -galactosidase (scFv13)[155]. Like PhoA, scFv13 folding is dependent on the formation of two intradomain disulfide bonds, one in the VH and one in the VL domain. In the absence of these bonds, the scFv is only expressed at very low levels[155]. Similar to what was seen above, DHB4 and DHA cells expressing ssDsbA-scFv13-Bla exhibited distinctly different Amp resistance phenotypes (Fig. 2.2c). The MBC was 8-fold greater for DHB4 cells compared to DHA cells expressing this construct (Table 2). Western blot analysis revealed significant accumulation of soluble ssDsbA-scFv13-Bla in the periplasmic fraction of DHB4 cells, while virtually no cross-reacting bands were detected in any fraction (cytoplasmic or periplasmic) isolated from DHA cells (Fig. 2.2d). This result confirmed that scFv13 folding and solubility was dependent on disulfide bond formation. The fact that scFv13-Bla accumulated predominantly in the periplasm provided additional support for our hypothesis that proteins targeted to the SRP pathway are efficiently partitioned in the periplasm. It is also noteworthy that, unlike the case of PhoA-Bla, we did not observe any proteolytic degradation of the scFv13-Bla fusion. We also tested scFv13-R4, a variant of scFv13 engineered previously to have greater solubility under reducing conditions[155]. In DHA cells, where the periplasm does not support disulfide bond formation, scFv13-R4 showed a two-fold increase in MBC over wild-type scFv13 (Table 2), consistent with the improved ability of scFv13-R4 to fold correctly in the absence of disulfide bonds[155]. Interestingly, expression of ssDsbA-scFv13-R4-Bla in DHB4 cells resulted in an 8-fold higher MBC relative to expression in DHA cells. This is likely due to the fact that oxidation of scFv13-R4 in the periplasm further increases the proteins stability by the addition of at least one disulfide bond[155]. Like its parent, scFv13-R4 accu-

Table 2.2: Amp resistance conferred by mammalian proteins

Protein ¹	Domain ²	Subcellular location	ssDsbA MBC ($\mu\text{g}/\text{ml}$) ³	ssTorA MBC ($\mu\text{g}/\text{ml}$) ³	STE ratio ⁴
CASP2 (Hs)	FL	cytoplasm	50	50	0.00
CCND2 (Hs)	FL	cytoplasm	100	9	0.12
CD44 (Hs)	FL	extracellular	6	6	0.13
CDK4 (Hs)	FL	cytoplasm	50	12	0.25
CDKN1B (Hs)	FL	cytoplasm	400	50	0.10
Efna1 (Mm)	EC	extracellular	200	25	0.36
Efnb2 (Mm)	EC1	extracellular	200	9	0.06
Efnb2 (Mm)	EC2	extracellular	800	9	0.32
Epha2 (Mm)	LB	extracellular	6	6	0.03
Ephb2 (Mm)	SAM	cytoplasm	800	100	0.16
Ephb2 (Mm)	TK	cytoplasm	100	3	0.07
FOS (Hs)	FL	nuclear	100	25	0.01
GATA2 (Hs)	FL	nuclear	3	500	0.09
GFP (Av)	FL	cytoplasm	200	400	0.69
GRB2 (Hs)	FL	cytoplasm	800	30	0.31
HRAS (Hs)	FL	cytoplasm	3	25	0.52
JUN (Hs)	FL	nuclear	400	25	0.05
MAD (Hs)	FL	nuclear	6	25	0.52
Mdm2 (Mm)	p53-bd	nuclear	400	12	0.42
MMPI (Hs)	FL	extracellular	200	50	0.09
RAF1 (Hs)	Ras-bd	cytoplasm	200	400	0.52

mutated stably in cells with no apparent degradation (Fig. 2.2a).

Evidence for a universal assay of periplasmic protein folding.

We next tested whether changes in folding behavior unrelated to disulfide bond formation could be discriminated using this selection strategy. Specifically, we

¹Organism: Hs, Homo sapiens; Mm, Mus musculus; Av, Aequoria victoria

²Domain: FL, full-length; EC, extracellular; TK, tyrosine kinase; LB, ligand binding, SAM, sterile alpha motif; bd, binding domain

³Approximately 500 CFUs were plated overnight at 37°C

⁴Soluble versus total expression (STE) ratios were obtained for each protein by normalizing the soluble expression data by the total expression data reported in Tables 2 and 3 of Dyson et al. (2004). Values reported are the average over all 6x and 10x his-tagged constructs.

tested several proteins that are physiologically expressed in the *E. coli* cytoplasm (e.g., GST, MetF, MetK and TrxA) as well as GFP. We reasoned that some of these would fold correctly in the periplasm. Indeed, selective plating of cells expressing these constructs revealed that GST, TrxA and GFP are each correctly folded in the periplasm as evidenced by MBC values of 50, 100 and 200 $\mu\text{g}/\text{ml}$ Amp, respectively (Table 2). For comparison, expression of Bla fused to the maltose binding protein (MBP), a native periplasmic protein that is known to be extremely soluble in *E. coli*[111], resulted in an MBC of 250 $\mu\text{g}/\text{ml}$ Amp (Table 2). Each of these was also found to stably accumulate in cells with little to no proteolytic degradation (Fig. 3a). We further reasoned that other proteins such as MetF and MetK would misfold and/or aggregate outside of their native environment. This is because MetF folding involves the binding of a flavin cofactor, which would likely not be present in the periplasm[207]. Similarly, MetK has redox-sensitive Cys residues that could become mis-oxidized in the periplasm. In addition, both proteins are strictly dependent on the cytoplasmic chaperone GroEL for correct folding[115]. As expected, the MetF and MetK fusions conferred little to no resistance to cells plated on 100 $\mu\text{g}/\text{ml}$ Amp (Fig. 3b) and resulted in MBC values of only 25 and 6 $\mu\text{g}/\text{ml}$, respectively (Table 2). The weak Amp resistance conferred by MetF and MetK corresponded with the complete absence of each fusion protein following Western blot analysis (Fig. 2a), indicating that incorrect folding of each protein in the periplasm was followed by proteolytic degradation. To rule out the possibility that MetF and MetK are incapable of folding when fused to Bla, we expressed MetF-Bla and MetK-Bla that lacked the ssDsbA export signals. Each of these folded very efficiently in the cytoplasm as evidenced by the relatively high Bla activity measured in the lysates of cells expressing these constructs (Fig. 2b). For comparison, very little Bla

activity was measured in the lysates of cells expressing ssDsbA-MetF-Bla and ssDsbA-MetK-Bla. Thus, C-terminal fusions of Bla to MetF and MetK are indeed capable of correct folding when expressed in the cytoplasm, but become misfolded and inactive when localized in the periplasm. The observation that GFP folded correctly in the periplasm was surprising based on earlier findings that it could not attain a fluorescent conformation when localized in the periplasm via the Sec pathway[65, 72]. Indeed, cells expressing the ssDsbA-GFP-Bla fusion were non-fluorescent (Fig. 2c). However, the absence of cell fluorescence could not be attributed to global misfolding or aggregation of GFP in the periplasm because ssDsbA-GFP-Bla conferred a strong Amp-resistant phenotype to cells (Table 2 and Fig. 2c). For comparison, cells expressing GFP-Bla via the Tat export pathway resulted in highly fluorescent cells that were similarly Amp-resistant (Table 2 and Fig. 2c), consistent with earlier findings[71, 143]. From these data, it is clear that GFP is soluble in both the cytoplasmic and periplasmic compartments; hence, the inability of GFP to fluoresce in the periplasm is not attributable to its overall folding and stability in this compartment.

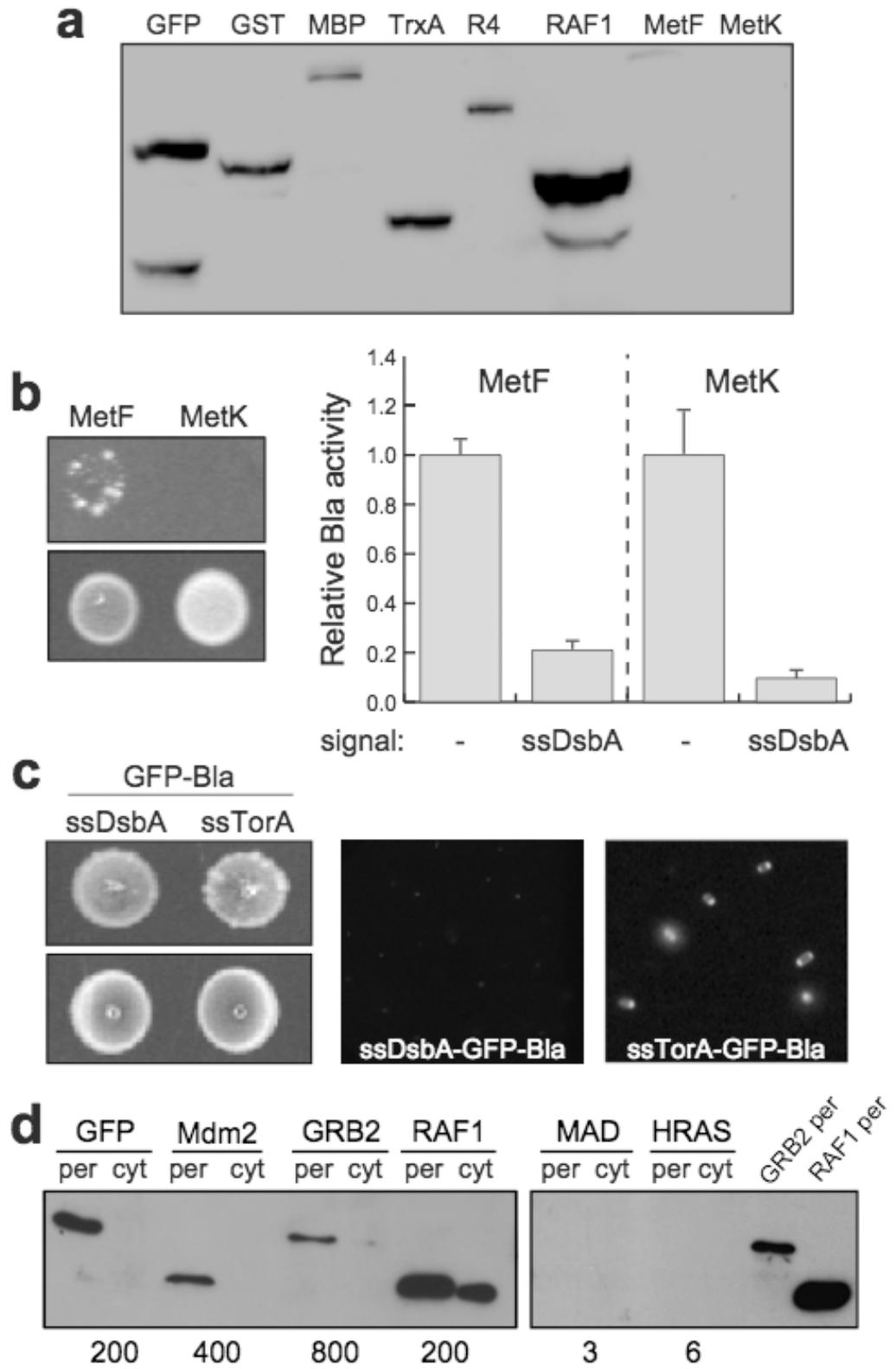


Figure 2.3: A universal folding reporter. (a) Western blot analysis of whole cell lysates generated from an equivalent number of DH5 α cells expressing the ssDsbA-POI-Bla constructs as indicated. Samples were blotted with Bla-specific antibodies. (b) An equivalent number of DH5 α cells expressing ssDsbA-MetF-Bla or ssDsbA-MetK-Bla were spot-plated on LB/agar containing either 50 μ g/ml Cm (- Amp) or 100 μ g/ml Amp (+ Amp). Nitrocefin hydrolysis activity of whole-cell lysates from cells expressing MetF-Bla and MetK-Bla with or without the ssDsbA signal peptide. Activity was measured as the initial velocity of absorbance change at 486 nm. Relative Bla activity was obtained by normalizing to the activity of the fusion without the signal peptide. Bla activity was measured in triplicate and error bars represent standard error of the mean. (c) DH5 α cells expressing ssDsbA-GFP-Bla or ssTorA-GFP-Bla were spot-plated as in (b). Fluorescence microscopy analysis (100X) of DH5 α cells expressing ssDsbA-GFP-Bla or ssTorA-GFP-Bla as indicated. (d) Western blot analysis of periplasmic (per) and cytoplasmic (cyt) fractions generated from an equivalent number of DH5 α cells expressing ssDsbA-POI-FLAG fusions lacking the Bla moiety for the clones indicated. Blots were probed by anti-FLAG antibodies. GroEL was used as a fractionation marker by probing with anti-GroEL antibodies (data not shown). GRB2 per and RAF1 per samples were included as positive loading controls for the blot on the right. MBC values (μ g/ml Amp) corresponding to each protein are indicated below the panels

To further test the limits of our selection strategy, we attempted to identify mammalian proteins that fold correctly in the periplasm. For this, a total of 20 mammalian proteins and protein domains were evaluated for SRP-mediated expression and selection. These proteins were of human or murine origin, and represented several diverse protein families with extracellular, cytoplasmic and nuclear cell locations (Table 2). We tested a collection of full-length and truncated proteins, which are described elsewhere[56], by cloning each into the pDMB selection vector as ssDsbA-POI-Bla fusions and determining the MBC for each. For 15 of the proteins tested, the MBC was 50 $\mu\text{g}/\text{ml}$ Amp (Table 2), suggesting that these were all relatively well folded and soluble in the periplasm. Of these, Efnb2(EC2), Ephb2(SAM) and GRB2 conferred a level of resistance (800 $\mu\text{g}/\text{ml}$ Amp) that was only 4-fold lower than the MBC measured for cells expressing ssDsbA-Bla with no target protein, indicating that these were extremely soluble when expressed in the periplasm. The resistance conferred by these 15 clones correlated with soluble expression of the fusion proteins, as exemplified by the RAF1-Bla fusion (Fig. 2a). Only 5 of the clones (CD44, Epha2(LB), GATA2, HRAS and MAD) exhibited MBCs of 6 $\mu\text{g}/\text{ml}$ Amp that were comparable to the MBC reported for plasmid-free wt MC4100 cells[142] and indicative of poor folding and solubility. It is informative to compare the MBC results obtained here for periplasmic folding versus the MBC values for cytoplasmic solubility determined previously using a Tat-mediated folding reporter that also employed TEM-1 Bla[143]. For some of the target proteins (e.g., CCND2, Efnb2(EC1), Efnb2(EC2), GRB2 and Mdm2-p53-bd), strong resistance was only conferred by periplasmic expression but not cytoplasmic expression, whereas GATA2, HRAS and MAD conferred greater resistance following cytoplasmic expression compared to periplasmic expression (Table 2). We also ob-

served that some of the proteins (e.g., EphB2(SAM), GFP, RAF1) were soluble in both the periplasmic and cytoplasmic compartments while certain others (e.g., CD44, Epha2(LB)) were not very soluble in either of these subcellular locations (Table 2).

To independently confirm soluble periplasmic expression, we expressed a subset of the mammalian clones as ssDsbA-POI-FLAG chimeras in which the C-terminal Bla moiety was replaced by a 9-residue FLAG epitope tag. Subcellular fractionation analysis of several representative positive clones (i.e., those conferring resistance to 50 $\mu\text{g}/\text{ml}$ Amp) including GRB2, Mdm2(p53-bd), RAF1 and GFP revealed that each of these accumulated in the soluble periplasmic fraction of wt cells (Fig. 2d). Conversely, we observed no soluble accumulation in either the periplasmic or cytoplasmic fractions for two negative clones, namely HRAS and MAD, that conferred resistance to 6 $\mu\text{g}/\text{ml}$ Amp. These results confirm that soluble expression in the periplasm correlates well with the resistance conferred by each of these clones.

cis- and trans-acting factors that affect protein folding in the periplasm.

We next determined whether the assay could report changes in folding robustness caused by cis-acting factors such as sequence mutations or trans-acting factors such as molecular chaperones. For these studies, we employed *E. coli* MBP and a collection of MBP variants that become kinetically trapped in off-pathway intermediates that are prone to aggregation[3]. One of these, a double mutant called MalE31 (G32D/I33P), reportedly forms inclusion bodies in the periplasmic space[18]. We cloned four different versions of MBP (wild-type (wt), G32D, I33P, MalE31) into pDMB and evaluated their ability to confer Amp resistance

to cells. Spot plating revealed that cells expressing wt MBP but not MalE31 were resistant to 200 $\mu\text{g}/\text{ml}$ Amp (Fig. 2.3a). Western blot analysis revealed that the resistance phenotypes conferred by ssDsbA-MBP-Bla and ssDsbA-MalE31-Bla correlated with the intracellular accumulation of each protein (Fig. 2b for MBP; Fig. 2.3b for MalE31). Plating of the same cells on LB lacking Amp showed no difference in growth phenotype for wt MBP versus MalE31 (Fig. 42.3a). The measured MBC for the wt MBP fusion was 42% greater than its MalE31 counterpart (Table2). It is noteworthy that while the MalE31 conferred significantly less Amp resistance than MBP, these Bla fusions were not completely inactive. This is in agreement with recent findings of Betton and coworkers who showed that despite its poor folding efficiency in the periplasm, a fusion of MalE31-Bla retained some catalytic activity[3]. Our previous findings showed that MalE31 expressed in the cytoplasm is somewhat soluble, albeit to a lesser extent than wt MBP[71]. It has been shown that the periplasmic chaperone FkpA can decrease aggregation of MalE31, while another periplasmic chaperone SurA did not affect MalE31 aggregation[4]. In addition, a periplasmic chaperone called Skp is well known for its ability to interact with a broad range of substrates[164] and has been used to improve expression of phage-displayed proteins[22]. Based on these data, we reasoned that a chaperone-mediated decrease in MalE31 aggregation would lead to increased Bla activity in our assay. To test this, cells co-expressing FkpA, SurA, or Skp with ssDsbA-MalE31-Bla were plated on Amp. FkpA and Skp co-expression with ssDsbA-MalE31-Bla resulted in a measurable increase in Amp resistance compared to the control case where ssDsbA-MalE31-Bla was expressed alone (Fig. 2.3b). In contrast, SurA co-expression conferred no measurable growth difference compared to the control (Fig. 2.3b). Cells grown on control plates lacking Amp showed no differences in growth pheno-

type (data not shown). Western blot analysis of periplasmic fractions confirmed that co-expression of FkpA and Skp, but not SurA, increased the solubility of ssDsbA-MalE31-Bla (Fig. 2.3c). Finally, to show that improved Amp resistance resulted from an increase in MalE31 solubility and not by improved activity of the Bla domain itself, we co-expressed each of the periplasmic chaperones with ssDsbA-Bla and found that Bla activity did not increase in the presence of any of the periplasmic chaperones (Fig.2.3d). In fact, co-expression of FkpA caused a decrease in Bla activity as evidenced by lack of cell growth on 200 $\mu\text{g}/\text{ml}$ Amp. Interestingly, we found that co-expression of SurA increased the solubility of the wt MBP fusion (Fig. 2.3b) suggesting that SurA promotes folding of wt MBP but not MalE31.

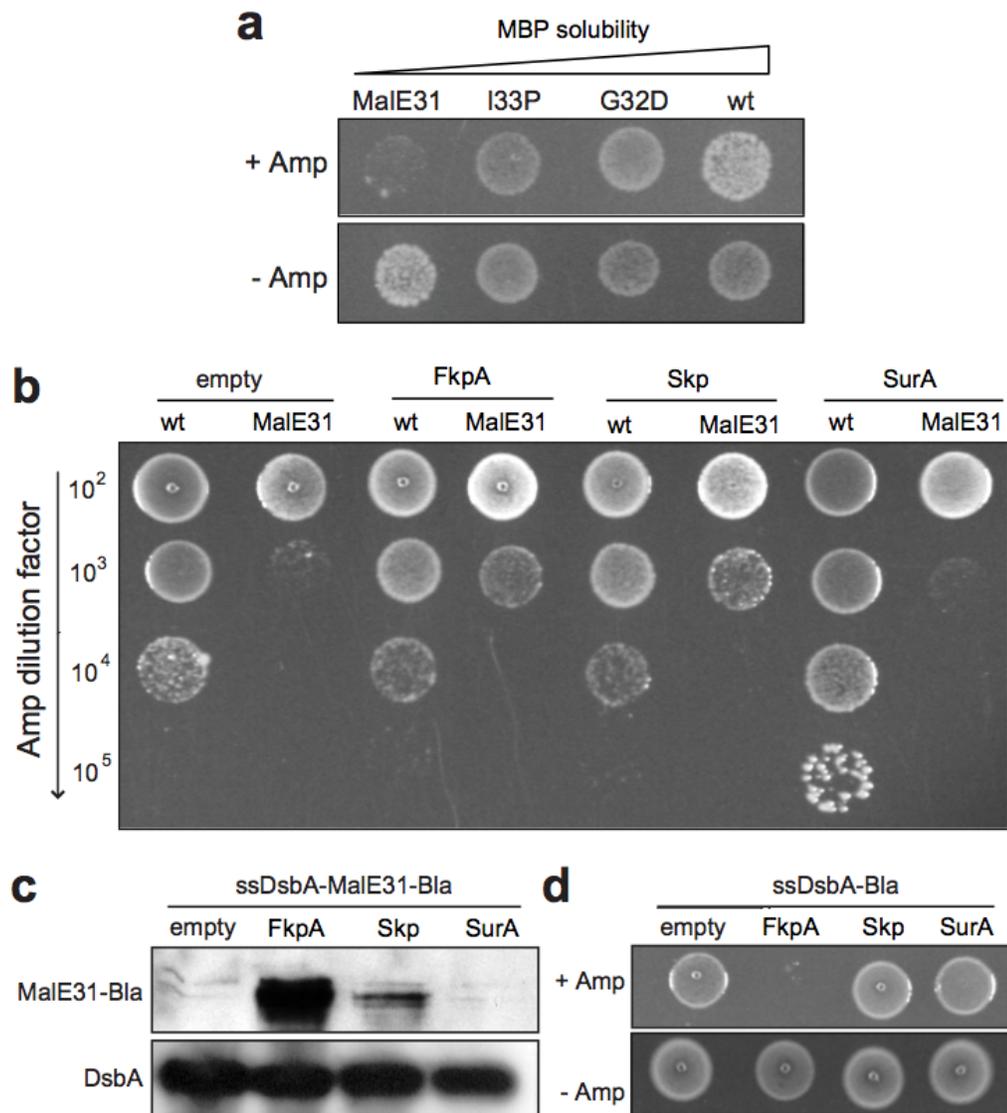
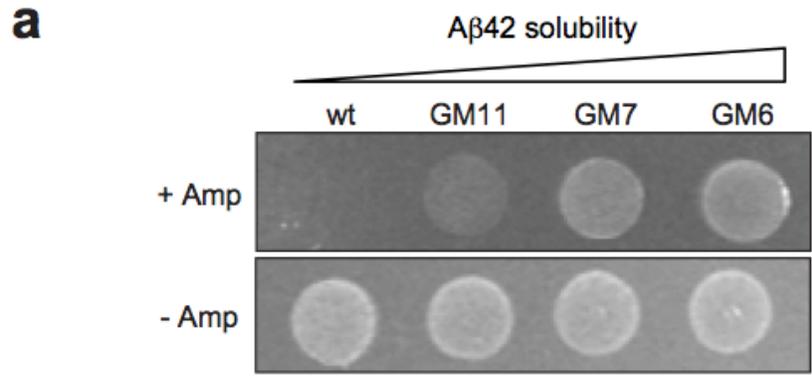


Figure 2.4: Effect of cis- and trans-acting factors on periplasmic protein folding. (a) DH5 α cells expressing wild-type (wt) MBP and its variants MalE31, I33P, G32D as sandwich fusions between ssDsbA and Bla. An equivalent number of cells were spot-plated on 100 μ g/ml Amp (+ Amp) or 50 μ g/ml Cm (- Amp). Open triangle indicates increasing solubility of MBP. (b) An equivalent number of DH5 α cells co-expressing periplasmic chaperones FkpA, Skp, and SurA (or with empty pBAD18-Kan) along with wild-type (wt) MBP or MalE31 fusions

were spot plated at various dilution factors (as indicated to left) on LB/agar supplemented with 200 $\mu\text{g}/\text{ml}$ Amp and 50 $\mu\text{g}/\text{ml}$ Kan. (c) Western blot analysis of periplasmic fractions from cells co-expressing FkpA, Skp, and SurA chaperones as indicated along with ssDsbA-MalE31-Bla. Samples were probed with Bla-specific antibodies. Detection of DsbA using anti-DsbA serum was performed as a loading control. (d) An equivalent number of cells co-expressing FkpA, Skp, and SurA chaperones as indicated along with ssDsbA-Bla were spot-plated on 50 $\mu\text{g}/\text{ml}$ Kan and either 800 $\mu\text{g}/\text{ml}$ Amp (+ Amp) or 20 $\mu\text{g}/\text{ml}$ Cm (- Amp).

Enhancing the solubility of aggregation-prone proteins.

To demonstrate the utility of this system, we next attempted to isolate solubility-enhanced protein variants in the periplasm using a directed evolution approach. For this purpose, we inserted the 42-residue amyloid- β 42 peptide ($A\beta$ 42) in the POI position of our folding reporter. $A\beta$ 42 is the primary constituent in dense amyloid fibrils that accumulate in the brains of patients with Alzheimers disease[88]. As in humans, $A\beta$ 42 aggregates extensively when expressed in *E. coli*[71, 253]. We and others have isolated variants of $A\beta$ 42 that were significantly more soluble in the bacterial cytoplasm compared to wt $A\beta$ 42[253, 71]. When expressed in the periplasm as a ssDsbA- $A\beta$ 42-Bla fusion, these solubility-enhanced variants conferred increased Amp resistance to cells relative to wt $A\beta$ 42 (Fig. 2.4a). The most soluble variant, clone GM6[253], displayed an 8-fold higher MBC relative to wt $A\beta$ 42 (Table 2).



c

wt DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA

GM6 -----S-----P-----

A2 -----T-----

A4 E-----

B9 -----D-----

B12 -----A-----

H2 -----P-----

Figure 2.5: Directed evolution of the $A\beta_{42}$ peptide. (a) Spot plating of an equivalent number of cells expressing fusions containing wild-type $A\beta_{42}$ and solubility-enhanced variants GM11, GM7 and GM6 on 100 $\mu\text{g}/\text{ml}$ Amp (+ Amp) or 50 $\mu\text{g}/\text{ml}$ Cm (- Amp). The variants are derivatives of $A\beta_{42}$ with the following mutations: GM11 (H6Q/V12A/V24A/I32M/V36G), GM7 (V12A/I32T/L34P) and GM6 (F19S/L34P). (b) Spot plating of an equivalent number cells expressing evolved variants of $A\beta_{42}$ with greater solubility than wild-type $A\beta_{42}$ at various dilution factors from overnight cultures. For comparison, wild-type $A\beta_{42}$ and soluble variant GM6 are shown. (c) Amino acid sequences of evolved $A\beta_{42}$ peptides, along with $A\beta_{42}$ and GM6 for comparison.

Encouraged by the observation that growth selection could easily distinguish more soluble $A\beta_{42}$ variants, we next attempted to isolate solubility-enhanced $A\beta_{42}$ variants from a combinatorial library of $A\beta_{42}$ sequences. An error-prone library of $A\beta_{42}$ in pDMB was created that contained approximately 5×10^4 members. Plating of approximately 10^4 cells from this library on 25 $\mu\text{g}/\text{mL}$ Amp resulted in 5 clones exhibiting an Amp-resistant phenotype above wt). Each of these variants showed at least a 2-fold improvement in MBC over wt $A\beta_{42}$, with clone H2 exhibiting a 4-fold increase in MBC compared to wt (Table 2). Interestingly, clone H2 contains the L34P mutation (Fig. 2.4c), which is known to render $A\beta_{42}$ less prone to aggregation[253]. The other isolated clones had point mutations located primarily within hydrophobic stretches in the central and C-terminal parts of $A\beta_{42}$ that have been implicated by others as key determinants in aggregation and fibril formation[253, 116]: A2 (M35T), B9 (V40D), B12 (V37A) (Fig. 2.3c). It is noteworthy that this improvement in $A\beta_{42}$ solubility only required a single round of mutagenesis and selection, confirming the robustness of our assay for searching large volumes of sequence space for proteins with altered structural characteristics.

Discussion

To date, several reporters of protein folding have been developed in *E. coli*[102, 156, 240, 250, 34]. Such methods include: transcriptional fusion reporters that signal the cells genetic response to misfolding and aggregation of the POI[130], and translational fusion reporters between the POI and a peptide (e.g., tetracysteine motif binding site[102]), protein (e.g., chloramphenicol acetyltransferase[156], GFP[240]) or split protein fragment (e.g., [250], GFP[34])

whose activity is modulated by the folding behavior of the POI. Along similar lines, we recently developed a protein solubility reporter that exploits the authentic protein folding quality control mechanism of the *E. coli* Tat pathway[71]. Surprisingly, there are currently no reports of an analogous reporter system for the periplasmic folding environment despite the significance of this compartment for the expression and engineering of heterologous proteins[11, 219, 80]. This is due in part to the fact that GFP, one of the most useful cellular reporter proteins that has been used to successfully assay cytoplasmic folding[240], fails to reach a fluorescent conformation when routed to the periplasm via the Sec export mechanism[65]. To remedy this, we developed a genetic reporter system comprising a protein of interest fused between an N-terminal co-translational export signal (ssDsbA) and a C-terminal selectable marker (Bla) that enables intimate coupling between periplasmic protein folding and antibiotic resistance. It should be noted that Bla gene fusions have been used previously in *E. coli* to: (1) isolate genes encoding exported and membrane proteins of prokaryotic[209] and eukaryotic[223] origin; (2) increase secretion efficiency of exported and membrane proteins[27]; and (3) dissect membrane protein topology and assembly[57]. However, no studies have used Bla for reporting the folding robustness of target proteins in the periplasmic space. Another unique aspect of our system is its use of the co-translational SRP pathway such that all folding events are relegated exclusively to the domain of the periplasm. One concern that arises when using Bla fusions is that proteolytic degradation can release active Bla. Indeed, our experiments revealed some degradation of the PhoA-Bla fusion. However, the degradation does not appear to have released functional Bla because the MBC of these cells (100 $\mu\text{g}/\text{ml}$) was notably lower than that seen for other cases where the ssDsbA-POI-Bla fusion was not prote-

olyzed (e.g., MBP-Bla, scFv13-Bla, scFv13-R4-Bla, etc.). Even though we cannot rule out the possibility in the future that certain protein fusions may be degraded to release functional Bla, we see no evidence in the cases that were tested here. In fact, most of the other Bla fusions tested in this study experienced little to no proteolytic degradation. Further, while the correctly folded ssDsbA-PhoA-Bla fusion was not completely stable, the misfolded counterpart was even less stable (i.e., completely degraded) as evidenced by the total absence of cross-reacting bands on the Western blot. Thus, the assay was still able to reliably report the folding robustness of the PhoA protein. It turns out that this was not a unique phenomenon as many of the incorrectly folded proteins (e.g., MetF, MetK) were efficiently degraded. In light of these results, we contend that any Bla activity that might arise due to fusion instability should not impact the usefulness of the assay provided that proper negative controls are applied, as was done here. Since in many cases protein expression in the periplasm is often advantageous compared to expression in the cytoplasm[11, 219, 80], this assay should be useful in the development and optimization of numerous biotechnological applications. For instance, we demonstrated that the assay is useful for the isolation of solubility-enhanced variants from a combinatorial library of protein sequences. These results indicate that this method could be used as a platform for (1) engineering proteins with superior solubility, i.e. superfolder proteins[178] and (2) pre-selection of large combinatorial libraries to eliminate incorrectly folded proteins since soluble structure is a prerequisite for function. In the future, it might be possible to use this system in combination with our recently reported Tat-based cytoplasmic folding reporter[71], that also relies on Bla activity, to comparatively and comprehensively explore the intracellular protein folding landscape. It might also be possible to use this selection in

conjunction with fluorescence activated cell sorting (FACS) to isolate GFP variants that fold and function in the periplasm. In addition to sequence-related determinants of protein folding, we have also shown that this assay can be used to probe extrinsic factors (e.g., molecular chaperones) that contribute to protein solubility in the periplasm. The ability to characterize or discover factors that affect folding in the periplasm and are orthogonal to protein sequence is intriguing. A more comprehensive model of the periplasmic space could have great implications for expression of proteins whose primary structure is constrained for reasons of, for example, therapeutic activity and immunogenicity. Besides chaperones, this assay can also be used to analyze how periplasmic post-translational modifications such as disulfide bond formation, as demonstrated here, or even *N*-linked glycosylation[220] affect folding and solubility. The stability conferred by disulfide bonds is often paramount to achieving soluble expression of many therapeutic proteins in *E. coli*, many of which are currently produced in *E. coli* by inclusion body formation and re-folding procedures due to disulfide-bond-related instability in the reducing cytoplasm. It should be noted that for those proteins whose proper folding depends on post-translational processing, the periplasmic folding reporter becomes a genetic reporter of that post-translational processing pathway. Finally, the fact that our assay can be used to evaluate the contribution of these processes to protein folding should help illuminate the various ways in which disulfide bonds and glycosylation reactions, or even their interplay[21], affects the structure, function and stability of proteins.

Materials and Methods Strains and plasmids. DHB4 *E. coli* cells (F' *lacIq* *pro/λ-ΔlacX74 galE galK thi rpsL phoR ΔphoA(PvuII) ΔmalF3*) or an isogenic derivative of DHB4, namely DHA, that carries the *dsbA::kan* allele[53], were

used for experiments where cellular redox state was investigated. All other experiments were performed in DH5 α or MC4100 *E. coli* cells as indicated. Cloning was performed using standard molecular biological techniques and protocols[197]. Plasmid pDMB was constructed by inserting DNA for the DsbA signal peptide (ssDsbA; DNA nucleotides 1-57 of the *E. coli* dsbA gene) between SacI and XbaI sites of pTrc99A-Cm[71]. Next, the gene encoding TEM-1 Bla was inserted between the BamHI and HindIII sites. Finally, genes encoding different POIs were inserted between the XbaI and Sall or BamHI sites resulting in a sandwich fusion between ssDsbA and Bla. The resulting plasmids contained very short linker sequences, Ser-Arg and Val-Asp-Gly-Ser, at the junctions of ssDsbA-POI and POI-Bla (see Fig. 2.3a), corresponding to the translated portions of the XbaI and Sall-BamHI sites, respectively. The POIs included: MetF, MetK, GST, PhoA and TrxA, all of which were amplified from the *E. coli* genome using colony PCR; maltose binding proteins MaleE, MaleE31, MaleE-G32D and MaleE-I33P[18] and single-chain antibodies scFv13 and scFv13-R4[155], which were kindly provided by J.-M. Betton; wild-type A β 42 peptide and solubility-enhanced A β 42 variants GM7, GM11, and GM6[253], which were kindly provided by M.H. Hecht; and GFP, which was PCR-amplified from pTMB-GFP[71]. Plasmid pTMB is identical to pDMB except that it contains the signal peptide of *E. coli* trimethylamine N-oxide reductase (ssTorA; DNA bases 1-126 of the *E. coli* torA gene) between SacI and XbaI sites. Genes encoding the periplasmic chaperones SurA, FkpA, and Skp were PCR-amplified from *E. coli* genomic DNA and inserted between the NcoI and Sall sites of pBAD18-Kan[87]. For mammalian protein expression (and GFP), a vector called pDSALK was created. This plasmid was created from a kanamycin-resistant version of pSALect[143] and contains the DsbA signal sequence between the NotI and XbaI restriction

sites. Mammalian proteins were cloned between the XbaI and either Sall or BamHI sites of pDSALK, then transformed into MC4100 cells for plating on Amp. For Western blot analysis of mammalian proteins expressed without the Bla moiety, proteins or protein domains with N-terminal ssDsbA were PCR-amplified from pDSALK constructs and cloned between the NcoI and HindIII sites of pTrc99A. A FLAG affinity tag was added to the C-termini of each protein or protein domain by PCR. Sequences of all plasmids constructed in this study were confirmed by DNA sequencing. Expression of fusion proteins and cell growth assays. Cells carrying a folding reporter plasmid were grown overnight at 37°C in LB medium containing 50 µg/mL chloramphenicol (Cm). Screening of cells on LB agar was performed by first normalizing overnight cultures by OD600 and then spotting 5 µL of serially-diluted (10-105-fold) cells on LB agar plates containing 100 µg/mL Amp or 20 µg/mL Cm. LB agar plates used to analyze co-expression of periplasmic chaperones were supplemented with 50 µg/mL kanamycin (Kan), along with either 20 µg/mL Cm or 200 µg/mL Amp and either 0.2% arabinose or 0.2% glucose. The trc promoter allows for leaky expression and therefore no IPTG was used to induce cultures. In all cases, the plates were incubated for 16 h at 37°C and then imaged using a Chemi-Doc System (BioRad). For MBC/MIC determination, approximately 200 colony forming units (CFUs) of each clone were plated on LB agar plates containing 0, 3, 6, 12, 25, 50, 100, 200, 400, 800, or 1600 µg/mL Amp or 20 µg/mL Cm. The MBC was determined as the minimum Amp concentration at which no colonies appeared on the plates. Minimum inhibitory concentration (MIC) was determined as the minimum concentration of Amp on which colony size or number of colonies was significantly smaller than control. Protein analysis. Cells were grown overnight at 37°C in flasks containing 50 ml LB media with appropri-

ate antibiotics. As above for the plating experiments, no IPTG was used to induce cultures because the *trc* promoter allows for sufficient leaky expression. Subcellular fractionation using the ice-cold osmotic shock procedure[53] was performed on an equivalent number of cells to generate soluble cytoplasmic and periplasmic fractions. Cytoplasmic fractions were obtained by sonication of resuspended spheroplasts following release of periplasmic proteins. Western blotting of these fractions was performed as previously described[53] using either 10 $\mu\text{g}/\text{mL}$ anti- β -lactamase, anti-FLAG M2 (Stratagene) at 1:500 dilution, anti-alkaline phosphatase antibodies (Sigma) at 1:20,000 dilution, or anti-DsbA serum diluted 1:5,000 (kindly provided by Dr. James Bardwell) as the primary antibody and anti-mouse or anti-rabbit horseradish peroxidase conjugate diluted 1:2,500 (Promega, Madison, WI) as the secondary antibody. Bands were visualized via chemiluminescent substrate (Bio-Rad) on Kodak film. The quality of all fractionations was determined by immunodetection of the cytoplasmic GroEL protein[53] or the periplasmic protein DsbA. Fractions from GFP-expressing cells were assayed for fluorescence by loading 100- μL portions into 96-well plates and quantifying the GFP activity (ex: 488 nm; em: 509 nm) using a microplate reader (Synergy HT, BioTek Instruments). Finally, soluble fractions were assayed for Bla activity based on nitrocefin (50 μM) hydrolysis in 96-well format as described[78]. All fluorescence and Bla activity measurements were performed in triplicate. Fluorescence microscopy. Cells expressing ssDsbA-GFP-Bla and ssTorA-GFP-Bla were visualized as described previously[121] using a Zeiss Axioskop 40 fluorescent microscope with Spotflex color digital camera and filter sets for GFP (485 nm for excitation and 505 nm for emission) and rhodamine (540 nm for excitation and 600 nm for emission). Library creation and selection of clones. A library of *A β 42* sequences was created according to

Fisher et al.[71], except that the plasmid backbone was pDMB. Briefly, error-prone PCR was performed on the gene encoding the A β 42 peptide. The gene library was cloned between the XbaI and SalI sites of pDMB and estimated to contain 50,000 members. Selection was performed by plating 2,000 CFUs per plate on 25 μ g/mL Amp. To eliminate false positives (e.g., small in-frame fragments that confer higher-than-expected resistance to Amp), clones growing on 25 μ g/mL Amp were inoculated in 96-well cultures and replica spot-plated as above at 10³ dilution on 25 and 100 μ g/mL Amp. Only those clones that grew on 25 μ g/mL Amp but failed to grow on 100 μ g/mL Amp were sequenced and characterized by MBC determination.

CHAPTER 3

AN *IN VIVO* ASSAY FOR BACTERIAL *N-LINKED* GLYCOSYLATION

Introduction

Protein glycosylation affects many protein properties including but not limited to folding, solubility, pharmacokinetic activity, and receptor binding. Glycosylation has been shown in many cases to enhance folding of glycoproteins [103]. In fact, it has been speculated that the addition of N-glycans has been an essential factor that has allowed eukaryotes to express high amounts of complex proteins that are well-folded [93]. Many proteins that are naturally glycosylated will express poorly or not at all in the absence of glycosylation [251].

Glycosylation was long thought to be exclusive to eukaryotic organisms. However, *N-linked* glycosylation was recently discovered in the pathogenic bacterium *Campylobacter jejuni* [221]. Glycosylation has been found to be essential for pathogenicity in this organism, but not overall cell viability [112]. In 2006, each gene in the 17-kb *pgl* (protein glycosylation) genetic locus was individually characterized by Kelly and colleagues [114]. The locus consists of 13 genes, most of the products of which participate in glycosylation. The process proceeds via five general steps: (1) saccharide and sugar-nucleotide biosynthesis in the cytoplasm, (2) attachment of the base sugar (bacillosamine) to lipid precursors on the membrane, (3) transfer of sugars to the glycan on the membrane, (4) "flipping" of the glycan from the cytoplasmic surface to the periplasmic surface of the inner membrane, and (5) transfer of the oligosaccharide to the asparagine residue at the glycosylation site of the protein. The consensus sequence for *C. jejuni* *N-linked* glycosylation has been determined to be

D/E-X₁-N-X₂-S/T, where X₁ and X₂ are any amino acid except proline [129]. In addition, the structure of the heptasaccharide was revealed by NMR spectroscopy to be GalNAc- α 1,4-GalNAc- α 1,4-[Glc β 1,3]GalNAc- α 1,4-GalNAc- α 1,4-GalNAc- α 1,3-Bac- β 1,N-Asn, where Bac is bacillosamine, 2,4-diacetamido-2,4,6-trideoxyglucose. While the enzymes involved in this pathway have been relatively well-characterized, the mechanism of involvement of N-linked glycosylation in infection is not currently well-understood.

Not long after, the *pgl* locus of *C. jejuni* was successfully transferred into *E. coli*, and it was shown that *E. coli* cells were capable of glycosylating proteins using the glycosylation machinery of *C. jejuni* [239]. Since the transfer of *C. jejuni* glycosylation to *E. coli*, several attempts have been made to engineer the process. First, Schwarz et al., swapped the oligosaccharyltransferase (OST) PglB from the *C. jejuni* system for a homologous protein from the related organism *Campylobacter lari*. The *C. lari* OST was found to have relaxed substrate specificity for the consensus sequence compared to the *C. jejuni* OST. However, the *C. lari* OST was not able to glycosylate a eukaryotic protein at its native consensus site, N-X-S/T [202]. In another work, Schwarz et al. also showed that a modified *pgl* locus, lacking the bacillosamine biosynthesis and transferase genes *pglCDEF* and glycosyltransferase *pglI*, synthesized a modified glycan that utilizes N-acetylglucosamine in place of bacillosamine and lacks the branched glucose motif. Proteins with this motif were then subjected to *in vitro* enzymatic processing to produce protein bearing eukaryotic-like glycans [201].

The only high-throughput screening method currently available for analysis of glycosylation is the recently reported glycophage display system, published almost simultaneously by Çelik et al. [36] and Dürr et al. [55]. Glycophage dis-

play attaches a glycoprotein to the pIII protein of a bacteriophage, which can then be panned for affinity to a glycan. The phagemid encoding the glycoprotein can then be recovered from bound phage particles. This technique shows promise for engineering of glycosylation in *E. coli*, but suffers from limitations due to its dependence on the affinity of a glycan-binding protein for the sugar group of interest, incorporation of the glycoprotein fusion into the phage particle, and the efficiency of glycosylation in *E. coli*, which is currently relatively low.

We recently developed a rapid protein folding assay for proteins in the periplasm of *E. coli* [153]. Since protein glycosylation occurs in the periplasm, we reasoned that adapting the protein folding assay for glycoproteins in *E. coli* would allow us to (1) study the effect of glycosylation on folding of various *C. jejuni* glycoproteins and (2) develop a genetic selection with the purpose of engineering the glycosylation pathway using the versatile *E. coli* as a host. Here we present the application of an assay for folding and solubility of glycoproteins and its application to analysis of glycosylation in *E. coli*.

We hypothesized that since glycosylation can affect the folding and solubility properties of proteins, a reporter of periplasmic protein folding could report changes in folding robustness effected by protein glycosylation. The assay is based on a tripartite fusion between: (1) an N-terminal signal peptide from *E. coli* DsbA (ssDsbA), which has previously been shown to direct proteins through the bacterial signal recognition particle (SRP)-dependent translocation pathway[199], (2) the glycoprotein of interest, and (3) a C-terminal fusion of mature TEM-1 β -lactamase (Bla). We have shown that the periplasmic folding assay couples folding robustness to an ampicillin resistance phenotype.

In this study, we analyzed 18 different reported glycoproteins from *C. jejuni*. In four cases we observed altered growth on ampicillin in the presence of functional glycosylation machinery. Thus, simple selection on β -lactam antibiotics such as ampicillin (Amp) enabled discrimination between glycosylated and aglycosylated proteins. We also inserted glycosylation sites into a bacterial protein the folding behavior of which has been shown to be affected by glycosylation *in vitro*. We were able to recapitulate this result in an *in vivo* assay.

Results

An *in vivo* assay for protein glycosylation

Peb3 is a glycoprotein adhesin from *C. jejuni* which has been hypothesized to be involved with the transport of 3-phosphoglycerate from the extracytoplasmic space [162]. It contains one glycosylation site at ¹¹⁸DSNIT¹²². As one of the first *N*-linked glycoproteins to be discovered in *C. jejuni*, its crystal structure with and without glycan has been solved [162]. In addition, a K135E mutation to wild-type Peb3 has been found to allow for more efficient glycosylation. While it was discovered that the overall structure of the protein did not differ significantly with the addition of the glycan, it was found that the glycosylated K135E variant possessed increased thermal stability of approximately 4.7°C over the aglycosylated protein.

Because this protein is well-characterized, we assayed its overall folding robustness using the previously reported folding assay. Briefly, we created a fu-

sion of the gene encoding Peb3 to mature TEM-1 Bla and directed the polypeptide for co-translational transport using the N-terminal signal sequence of *E. coli* DsbA. We then transformed this plasmid into CLM24 cells [67] containing either pACYC-pglKan or pACYC-pglmutKan. These vectors express the entire pgl operon on a plasmid, with the sole difference being the incorporation of two inactivating point mutations into the pglB gene in the case of pACYCpglmutKan. The cells were grown overnight and spot plated on varying levels of ampicillin to characterize the Bla activity. Figure 3b shows the results of this assay.

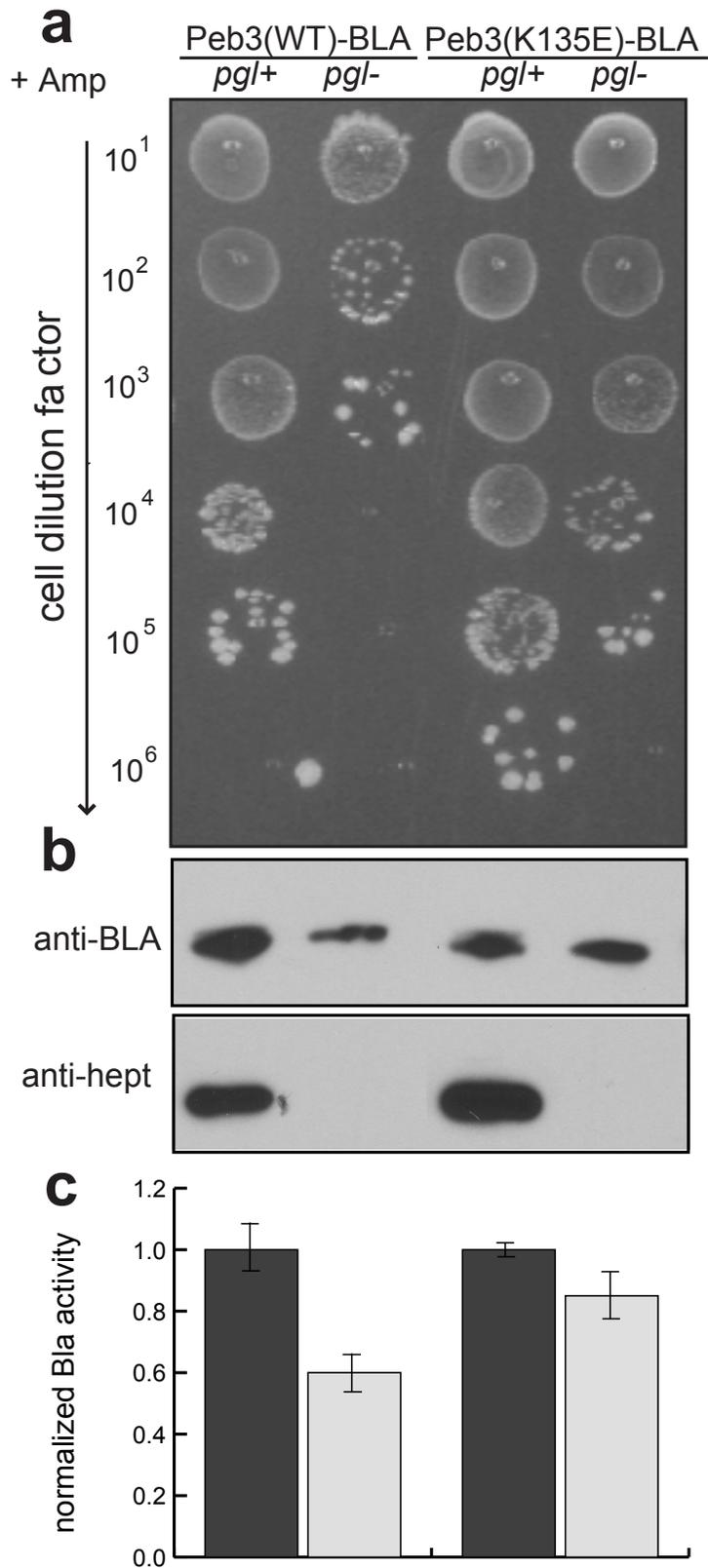


Figure 3.1: Glycosylation-dependent folding reported via SRP selection. (a) An equivalent number of CLM24 cells expressing ssDsbA-Peb3-Bla (WT or K135E) and either pACYC-pglKan (pgl+) or pACYC-pglmutKan (pgl-) were spot-plated on LB/agar containing 200 $\mu\text{g}/\text{ml}$ Amp and 50 $\mu\text{g}/\text{ml}$ Kan. (b) Western blot analysis of periplasmic fractions derived from CLM24 cells expressing the above constructs. Samples were blotted with Bla-specific antibodies (anti-BLA) or hR6 serum, which is specific to the *C. jejuni* heptasaccharide. (c) Nitrocefin hydrolysis activity of periplasmic fractions of above constructs. Activity was measured as the initial velocity of absorbance change at 486 nm. Specific Bla activity was obtained by normalizing to the total protein present in the periplasmic fraction. Bla activity was measured in triplicate and error bars represent standard error of the mean.

In both wild-type and K135E variants, Peb3 shows an apparent increased growth on ampicillin in the case of active glycosylation machinery. To confirm that the growth increase was due to increased Bla activity, we performed a Western blot analysis with an anti-Bla antibody (AbCam) (Figure 2.3b). Immunoblotting with hR6 antiserum, which is specific to the *C. jejuni* glycan, confirmed that Peb3 was glycosylated in both cases. Comparison of the bands in the hR6 (anti-heptasaccharide) blot also confirmed that glycosylation of K135E is more efficient than wild-type, though perhaps not as dramatically as the reported difference when expressed in *C. jejuni* [162]. To confirm the increase in Bla activity, we performed an *in vitro* Bla assay using the colorimetric substrate nitrocefin. The glycosylated case showed an increase in BLA activity of 67% for the wild-type and 20% for K135E. Thus, we were able to show that ampicillin resistance was correlated with protein glycosylation.

Scanning the *C. jejuni* periplasmic glycoproteome

To determine if any other proteins behaved similarly vis-à-vis glycosylation, we tested an array of putative periplasmic glycoproteins from *C. jejuni*. The proteins selected were confirmed to be glycosylated in an exhaustive glycoproteomic mass spectrometry study by Scott et al. [203]. In this study, which more than doubled the number of known *C. jejuni* protein glycosylation sites, the authors identified glycopeptides belonging to proteins localized on the inner membrane, periplasm, and outer membrane/extracellular space. Of the proteins identified, 28% are predicted to localize in the periplasm. We amplified 18 genes from *C. jejuni* genomic DNA and inserted them into the ssDsbA-POI-Bla fusion to assess their folding robustness in *E. coli* in the presence and absence of

Table 3.1: Single colony MIC and MBC measurements for *C. jejuni* proteins tested in this assay.

Protein	MIC/MBC in $\mu\text{g}/\text{mL}$ Amp (p _g l ⁺)	MIC/MBC in $\mu\text{g}/\text{mL}$ Amp (p _g l ⁻)
Cj0114	25/25	25/25
Cj0143c	50/50	50/50
Cj0200c	400/400	400/400
Cj0168	200/400	400/400
Cj0515	12/25	12/12
Cj0610c	100/100	200/400
Cj0648c	800/800	800/800
Cj0694	50/50	50/50
Cj0843c	12/25	12/25
Cj1345	50/50	50/50
Cj1444	200/400	200/200
Cj1496	800/800	400/800
Cj1621	12/12	12/12
1670c	25/50	25/50
CjaA	400/400	100/200
HisJ	25/50	25/50
PEB3 WT	800/800	400/400
PEB3 K135E	400/800	400/400

glycosylation.

Of the proteins tested, most showed no significant change in ampicillin resistance in the presence or absence of glycosylation (Table 3.1). One notable exception was the CjaA protein. CjaA, also known as Cj0982, is a putative cysteine transporter, a major surface antigen and vaccine candidate. In *C. jejuni*, CjaA is a lipoprotein that is located on both the inner membrane and cell surface [167]. However, recent studies have shown that when expressed in *E. coli*, CjaA primarily accumulates in the periplasm [254]. No crystal structure of the glycosylated form is available, however the aglycosylated form has been crystallized and its structure determined [167]. CjaA has one glycosylation site at ¹³⁷DSNIS¹⁴¹.

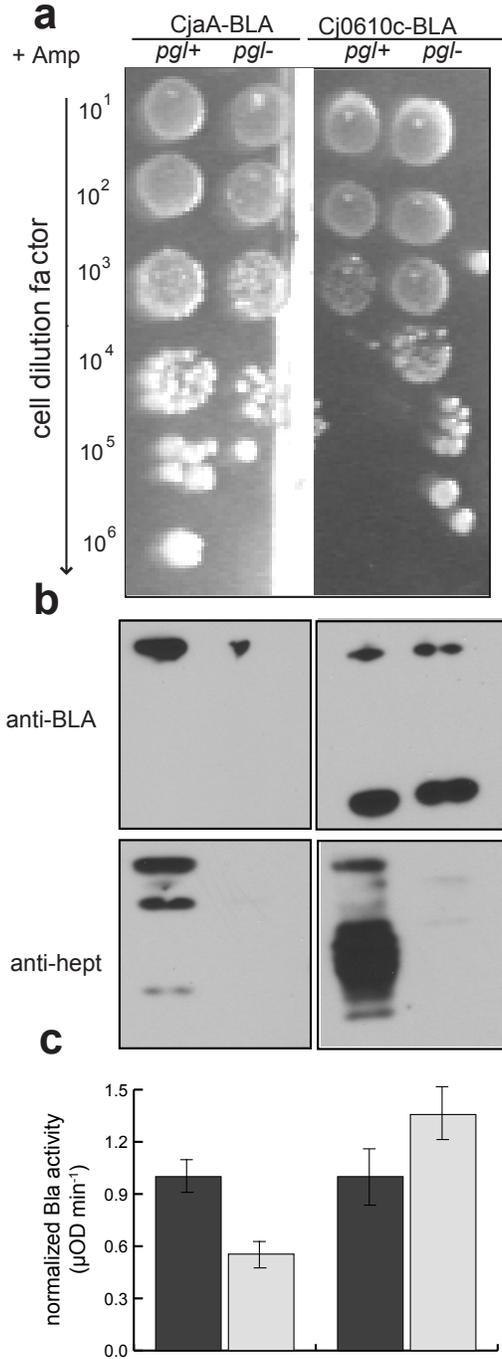


Figure 3.2: Glycosylation-dependent folding of CjaA and Cj0610c. (a) Spot plating as in Figure 3 on 200 $\mu\text{g}/\text{mL}$ Amp and 50 $\mu\text{g}/\text{mL}$ Kan (b) Western blot analysis as in Figure 3 (c) Nitrocefin hydrolysis activity as in Figure 3

When the ssDsbA-CjaA-BLA was plated on ampicillin, we again observed a positive correlation between glycosylation and ampicillin resistance (Fig This effect was confirmed by Western blotting with anti-Bla and hR6 serum). Nitrocefin assays confirmed that normalized Bla activity of periplasmic fractions expressing CjaA-BLA was approximately 80% higher in the glycosylated case (Figure 3.3).

Interestingly, not all proteins whose ampicillin-resistance phenotype was impacted by glycosylation showed increased growth in the glycosylated case. Spot plating of the fusion of protein Cj0610c-Bla actually results lower growth on ampicillin in the presence of functional glycosylation machinery (Figure 3.2). Cj0610c is a putative periplasmic protein with homology to SGNH hydrolase (as determined by sequence alignment by BLASTP, NCBI), a peptidoglycan o-acyltransferase. Since this protein is poorly-characterized, it is not immediately clear why glycosylation of this protein in a *C. jejuni* host would impact folding robustness in a negative way.

Glycosylation affects folding of a bacterial immunity protein

in vitro studies of the bacterial immunity protein im7 have recently elucidated several residues where introduction of a glycosylated Asn residue affected the overall ΔG of folding. Glycosylation at position 27 was determined to increase the overall stability of the protein by approximately 4 kJ/mol [41]. The constructs in question were created by protein ligation *in vitro*. To determine the effects of glycosylation on this protein *in vivo*, we sought to introduce the im7 protein into our glycosylation assay. The protein does not natively contain the glycosylation consensus sequence D/E-X₁-N-X₂-S/T, so engineered versions of

the protein were created by site-directed mutagenesis. The two mutants created encode either ²⁵DQNAT²⁹ or ²⁵ENNAT²⁹, where N is residue 27. DQNAT was chosen because of its efficiency of glycosylation in previous proteins [74], while ENNAT was chosen to maintain the closest proximity to the original protein sequence in this area (²⁵ENVAAT³⁰, where V is WT residue 27). Figure 3.3b shows the results of spot plating assays on the wild-type (aglycosylated), DQNAT and ENNAT versions of the im7 protein. In both the DQNAT and ENNAT cases, increased growth on ampicillin was observed in the cells with functional glycosylation machinery. Western blotting confirmed that the appropriate constructs were glycosylated. Some degradation of the constructs was observed, but this is not uncommon in the case of fusion proteins [153]

Discussion

In this study, we observed four proteins whose growth on ampicillin was coupled positively to protein glycosylation. This leads us to speculate on the feasibility of developing a genetic selection for glycosylation in *E. coli* based on growth on ampicillin. Further study and optimization of glycosylation conditions may prove fruitful in this area.

We report that *C. jejuni* proteins Peb3 and CjaA showed increased folding robustness when glycosylated. Many of the other *C. jejuni* proteins tested did not show a significant difference in ampicillin-resistance phenotype due to glycosylation (Table 3.1). To date, the impact of glycosylation on these proteins has only been studied in a few cases. For example, glycosylation of Cj0143c was found not to affect zinc transport or chick colonization [49]. Cj1496c function

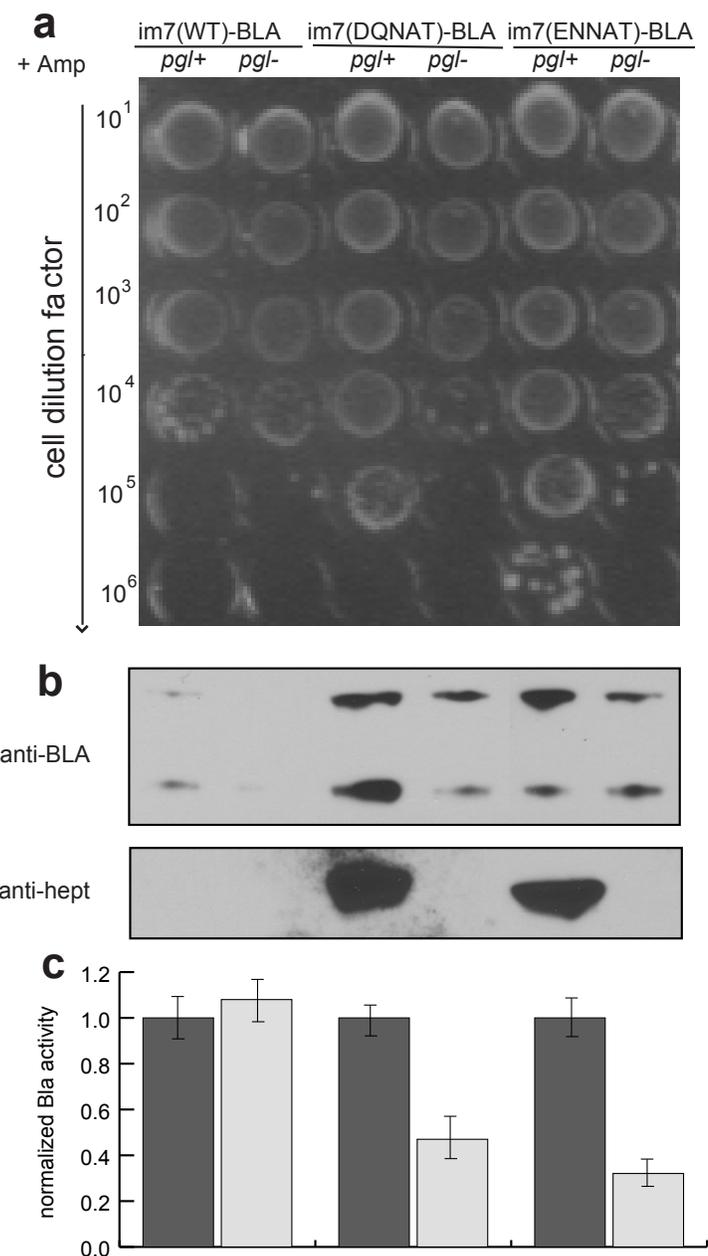


Figure 3.3: Glycosylation-dependent folding of im7 wild-type, DQNAT, and ENNAT constructs (a) Spot plating as in Figure 3 on 400 $\mu\text{g}/\text{mL}$ Amp and 50 $\mu\text{g}/\text{mL}$ Kan (b) Western blot analysis as in 3 (c) Nitrocefin hydrolysis activity as in 3

was also not affected by glycosylation [108].

We suggest the following explanations for a lack of difference in behavior between glycosylated and aglycosylated proteins: (1) Glycosylation in *E. coli* is relatively inefficient. The highest reported efficiency of glycosylation of model proteins in *E. coli* is roughly 40% of total extracted protein [176] and this was achieved after significant optimization of the process by an intensive proteomic study. Quantitatively measuring the efficiency of glycosylation is difficult, but it is likely safe to say that glycosylation is relatively inefficient. (2) It is possible that the glycan simply has no effect on overall protein folding. The contribution of the heptasaccharide may be minimal depending on its location in the protein, its ability to form hydrogen bonds, and the overall solubility of the protein. (3) The kinetics of translocation, folding, and glycosylation may be different in the reconstituted system in *E. coli* than in the native environment of *C. jejuni*. There are many examples in nature of glycans assisting in the protein folding process [94], since the initial oligosaccharyltransferase step in eukaryotes occurs co-translationally, though this phenomenon has not been thoroughly investigated in *C. jejuni*. It is certainly possible that overexpression of the proteins of interest on medium-high copy plasmids via cotranslational translocation (and in a different organism with a different folding milieu) is not optimal for observing the full impact of glycosylation on a protein's native folding pathway. (4) Glycosylation has been shown to increase thermal stability of proteins in many cases [5]. It is generally thought that this occurs by destabilization of the unfolded state of the protein, leading to an overall decrease in ΔG of folding. Cells in this study were grown at 30°C, but further study at a range of temperatures may help elucidate changes brought about by glycosylation.

On the other hand, the case of Cj0610c is interesting because it was found that glycosylation impacted folding robustness negatively. Glycosylation has been shown to negatively impact protein folding, specifically in the case of the im7 protein [41]. Most of the glycosylated asparagines that were inserted into im7 in the aforementioned study disrupted folding, while only one (at V27) was observed to improve folding. This property of Cj0610c or other negatively-impacted proteins could serve as the basis for high-throughput screens of inhibitors of glycosylation, which could lead to (1) greater understanding of the overall glycosylation pathway and (2) potential drug targets against infection by *C. jejuni*.

As the study of glycosylation in *E. coli* continues to evolve, it is clear that engineering of the glycosylation machinery itself and/or the strains that produce glycoproteins is necessary to make glycoprotein production in *E. coli* a viable alternative to mammalian cell culture. Therefore, tools which allow for high-throughput screening of the glycosylation status of a protein will undoubtedly help to make this goal achievable.

Materials and Methods

Strains and Plasmids

Cloning was performed using standard molecular biological techniques and protocols[197]. Genes from *C. jejuni* were amplified by PCR from genomic DNA of *Campylobacter jejuni* subsp. *jejuni* NCTC 11168. The RNaseA-S32D variant was amplified from pMIK81[128] (Markus Aebi). Variants of the im7

protein were amplified from constructs provided by James Bardwell and mutants were created by nested PCR. Plasmids used for this study were constructed using a modified version of vector pDMB, the construction of which has been shown previously [153]. Briefly, plasmid pDMB was constructed by inserting DNA for the DsbA signal peptide (ssDsbA; DNA nucleotides 1-57 of the *E. coli* dsbA gene) between SacI and XbaI sites of pTrc99A-Cm[71]. Next, the gene encoding mature TEM-1 Bla was inserted between the BamHI and HindIII sites. The pDMB vector was modified to contain two SfiI restriction sites flanking the protein of interest. This was accomplished by cloning the gene encoding Cj0200c between XbaI and Sall sites in pDMB using the primers 5'-aaaTCTAGAGGCCGGCTCGGCCTTgatagtctaaaacttgaagggtactattgcac-3' and 5'-ttgtcgacAAGGCCttcagGGCCataagctttttataacattgaatttcaatttttc-3'. Subsequent cloning of all genes used in this study occurred between the SfiI sites, except for *peb3* and variants, which were cloned between XbaI and Sall. The K135E variant of Peb3 was generated by site-directed mutagenesis using overlap extension PCR. Genes from *C. jejuni* were amplified starting after the encoded signal sequence, as determined by SignalP 3.0 [60]. Plasmids pACYC-pglKan and pACYC-pglmutKan were also kindly provided by Markus Aebi. Vector pDMB-POI and either pACYC-pglKan or pACYC-pglmutKan were co-transformed into strain CLM24 [67]. Sequences of all plasmids constructed in this study were confirmed by DNA sequencing.

Expression of fusion proteins and cell growth assays.

Cells carrying a folding reporter plasmid were grown overnight at 37°C in LB medium containing 50 µg/mL chloramphenicol (Cm). Screening of cells on LB

agar was performed by first normalizing overnight cultures by OD600 and then spotting 5 μ L of serially-diluted (10-105-fold) cells on LB agar plates containing 50 μ g/mL kanamycin and either 20 μ g/mL Cm or 0, 6, 12, 25, 50, 100, 200, 400, 800, or 1600 μ g/mL ampicillin. In all cases, the plates were incubated 16 h at 30°C and then imaged using a ChemiDoc System (BioRad). For MBC/MIC determination, approximately 200 colony forming units (CFUs) of each clone were plated on LB agar plates containing 0, 3, 6, 12, 25, 50, 100, 200, 400, 800, or 1600 μ g/mL Amp or 20 μ g/mL Cm. The minimum bacteriocidal concentration was determined as the minimum Amp concentration at which no colonies appeared on the plates. Minimum inhibitory concentration (MIC) was determined as the minimum concentration of Amp on which colony size or number of colonies was significantly smaller than control.

Protein analysis.

Overnight cultures of cells were inoculated into 10 mL LB with 40 μ g/mL Cm and 100 μ g/mL Kan and grown at 37°C to mid-log phase. They were then induced with 0.1 mM IPTG and transferred to 30°C for 6 hours. Subcellular fractionation using the ice-cold osmotic shock procedure[53] was performed on an equivalent number of cells to generate soluble periplasmic fractions. Western blotting of these fractions was performed as previously described[53] using either 10 μ g/mL anti- β -lactamase (AbCam), or hR6 antiserum (provided by Markus Aebi) as the primary antibody and anti-mouse or anti-rabbit horseradish peroxidase conjugate diluted 1:2,500 (Promega, Madison, WI) as the secondary antibody. Bands were visualized via chemiluminescent substrate (Bio-Rad) on Kodak film. Finally, soluble fractions were assayed for Bla activity

based on nitrocefin (50 μ M) hydrolysis in 96-well format as described[78]. All Bla activity measurements were performed in triplicate.

CHAPTER 4

**MODELING BACTERIAL N-LINKED GLYCOSYLATION USING AN
ADAPTED GENOME-SCALE MODEL OF *ESCHERICHIA COLI***

Protein glycosylation is an important post-translational modification that has been shown to affect protein folding, solubility, pharmacokinetic activity, and other protein properties. Many therapeutic proteins are glycoproteins and the glycosylation state of a protein has implications for its application downstream. Recently, N-linked glycosylation, once thought to be exclusive to eukaryotes, was discovered in a bacterium, *Campylobacter jejuni*. The subsequent functional transfer of the gene locus responsible for glycosylation to the popular biotechnology host *E. coli* has opened up new opportunities for glycoprotein production in this well-characterized, reliable workhorse organism.

In 2006, each gene in the 17-kb *pgl* (protein glycosylation) genetic locus was individually characterized by Kelly and colleagues [114]. The locus consists of 13 genes, most of the products of which participate in glycosylation. The process proceeds via five general steps (Figure 4.1) : (1) saccharide and sugar-nucleotide biosynthesis in the cytoplasm, (2) attachment of the base sugar (bacillosamine) to lipid precursors on the membrane, (3) transfer of sugars to the glycan on the membrane, (4) "flipping" of the glycan from the cytoplasmic surface to the periplasmic surface of the inner membrane, and (5) transfer of the oligosaccharide to the asparagine residue at the glycosylation site of the protein. The consensus sequence for *C. jejuni* N-linked glycosylation has been determined to be D/E-X₁-N-X₂-S/T, where X₁ and X₂ are any amino acid except proline [129]. In addition, the structure of the heptasaccharide was revealed by NMR spectroscopy to be GalNAc- α 1,4-GalNAc- α 1,4-[Glc β 1,3]GalNAc- α 1,4-GalNAc- α 1,4-

GalNAc- α 1,3-Bac- β 1,N-Asn, where Bac is bacillosamine, 2,4-diacetamido-2,4,6-trideoxyglucose.

Currently, production of glycoproteins in *E. coli* modified to carry the *C. jejuni* glycosylation locus suffers from two major limitations. First, overall efficiency of glycosylation is limited. The highest reported efficiency of glycosylation to date for a model protein is approximately 40 percent of total extracted protein [176]. Second, the glycan attached to proteins is markedly different from eukaryotic glycans, being primarily a linear chain of N-acetylgalactosamine (GalNAc) residues with one branched glucose as compared to a bi- or tri-antennary branch comprising a tri-mannose core followed by GlcNAc, galactose, and in humans, sialic acid. For glycoprotein production in *E. coli* to be feasible and economically viable, these challenges must be met.

Only a few models of N-linked glycosylation have been published to date [233, 131, 119]. Recently, a genome-scale model of *Pichia pastoris* was adapted to account for protein glycosylation [43]. These models focus on glycosylation in eukaryotic organisms and largely take advantage of the spatial organization of enzymes present in the Golgi. Flipping of individual nucleotide-sugar precursors into the Golgi, a major step in eukaryotic glycosylation, has also been considered. In *E. coli*, no such spatial organization exists and the glycan is flipped into the periplasm *en bloc* [114].

Genome-scale models allow reserachers to study metabolic processes on the scale of a complete organism[190]. For metabolic engineering applications, modeling enables examination and simulation of metabolism as a whole, enabling the detection of pathways and reactions whose contributions to production of a given end product may be non-intuitive [76, 2]. The most up-to-date

genome-scale stoichiometric model of *E. coli* is iAF1260 [66], a model which encompasses 1260 open reading frames. This model is unique not only in its magnitude of reactions, but in the incorporation of a periplasmic space in the *in silico* organism. Glycosylation occurs in *E. coli* in the periplasm, so the fact that this model allows us to closely approximate the environment and reactions of the periplasm is ideal.

In this study, we sought to determine the metabolic effect of the introduction of glycosylation machinery into *E. coli*. We added the reactions corresponding to the biosynthesis, flipping, and transfer of glycan to a target protein to the iAF1260 model. Using flux balance analysis, we measured the effect of simulated gene knockdown and simulated gene overexpression on the production of glycosylated protein. With this information, we assayed several knockout strains of *E. coli* for production of glycan and compared the results to our findings

Results and Discussion

Adaptation of a genome-scale model of *E. coli* for N-linked glycosylation of proteins

We simulated knockdown of individual genes in *E. coli* by constraining the corresponding fluxes to zero in the model and maximizing production of glycoprotein. To select for conditions that did not impair cell growth, during these simulations biomass was constrained to be 90% of the maximum possible biomass production under the simulation conditions. We performed this experiment under uptake conditions similar to growth in LB media. This simulated LB

Table 4.1: Reactions that lead to knockdown of protein glycosylation when knocked out.

Reaction	Class	% Knockdown of protein glycosylation
Undecaprenol pyrophosphate regeneration	LPS synthesis	100
Undecaprenol transfer to cytoplasmic inner membrane	LPS synthesis	100
UDP-glucose synthesis	Glycan biosynthesis	100
Glucose-6-phosphate to glucose-1-phosphate isomerization	Gluconeogenesis	100
Ribulose-5-phosphate to Ribose-5-phosphate isomerization	Pentose Phosphate Pathway	96
Glutamate amination to glutamine	Amino Acid Biosynthesis	79
CO ₂ export	Exchange	68
1,3-bisphosphoglycerate to glyceraldehyde-3-phosphate	Gluconeogenesis	50
3-phosphoglycerate to 1,3-bisphosphoglycerate	Gluconeogenesis	50
Fumarate hydrolysis to malate	TCA Cycle	7
Succinate oxidation to fumarate	TCA Cycle	6

media contained all 20 amino acids plus the requisite minimal media salts. Table 4 shows the top results of the knockout experiment, with the exception of the reactions corresponding to the *C. jejuni* *pgl* locus, all of which lead to total knockdown of glycosylation.

While some reactions make intuitive sense, e.g., recovery of the phosphorylated lipid carrier into the cytoplasm and ribose phosphate synthesis for creation of nucleotide-sugar donors, some are less obvious. For example, the involvement of the TCA cycle is interesting. We hypothesize that the products of the TCA cycle are contributing to gluconeogenesis via pyruvate carboxylation and synthesis of sugars.

Flux Coupling Analysis

To systematically determine which reactions were coupled to protein glycosylation, we implemented the Flux Coupling Finder (FCF) algorithm [31] on the modified genome-scale model. Briefly, this algorithm finds coupled fluxes by first constraining a target reaction to a maximum and individually minimizing each reaction, then constraining each reaction to the determined maximum and minimizing the target reaction.

When using protein glycosylation with the *C. jejuni* glycan as the target reaction, we found that 195 reactions were directionally coupled to the target flux. Among these reactions were many of the reactions found in the TCA cycle and gluconeogenesis (Figure 4.2). Interestingly, the flux coupling finder isolated isocitrate lyase as coupled to glycosylation. This reaction diverts flux from the TCA cycle to the glyoxylate shunt. In doing so, it bypasses the decarboxylation of α -ketoglutarate, conserving carbon. This allows for more carbon-efficient generation of malate and oxaloacetate, which can be used for sugar biosynthesis via gluconeogenesis.

In a recent, large-scale proteomic study of protein glycosylation in *E. coli*, isocitrate lyase was found to be upregulated in cells producing glycoprotein [176]. Overexpression of this gene resulted in a 3-fold increase in glycosylation efficiency. We also found that the glyoxylate shunt was important in glycosylation and that the glyoxylate cycle and gluconeogenesis are coupled to glycan synthesis when cells are grown on simulated LB medium. The reason for the inclusion of this pathway is likely that gluconeogenesis is required for sugar biosynthesis since it has been demonstrated that cells grown on LB use amino acids as a carbon source [206] and very little fermentable sugar is available for

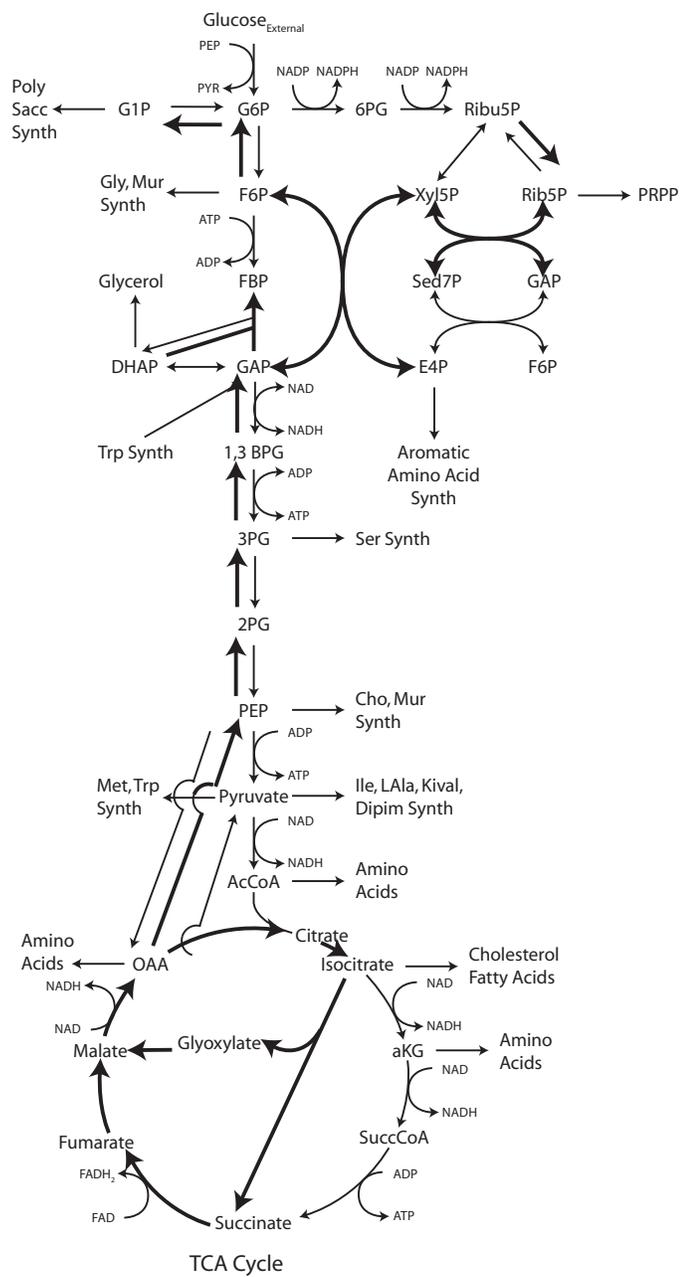


Figure 4.2: Reactions of the central carbon metabolism that are coupled to protein glycosylation. Arrows in bold indicate coupled reactions.

cell growth.

Reverse Knockdown Studies

We next sought to determine which reactions (if any) led to increased glycosylation when knocked down. Such reactions would likely be sources of diverted flux from the glycosylation pathway. The nature of the flux balance analysis algorithm makes this question difficult to answer since maximization of glycan will automatically zero out unconstrained competing reactions. Therefore, we solved the model for the maximum flux of each reaction with the constraint that the minimum biomass generation be 90% of the maximum biomass production given the uptake constraints (simulated LB). Next, each reaction was constrained to the maximum flux and glycosylation was maximized. This condition simulates overexpression of the gene corresponding to a particular reaction flux. Some of the reactions that decreased glycosylation when "overexpressed" are noted in Table 4.

Interestingly, many of the reactions that knock down glycosylation when overexpressed are the backward reactions of those found to decrease glycosylation when knocked out. For example, the interconversion of fumarate to succinate is obviously important, which underscores the overall importance of regulation of the TCA cycle in glycan biosynthesis and subsequent protein glycosylation.

Table 4.2: Reactions that lead to knockdown of protein glycosylation when over-expressed.

reaction	class	% knockdown of protein glycosylation
Citrate isomerization to cis-acotinate	TCA cycle	100
Ribulose-5-phosphate isomerization to arabinose-5-phosphate	Pentose Phosphate Pathway	100
Fumarate reduction to succinate	TCA cycle	75
Glyoxylate reduction to glycolate	Glyoxylate Metabolism	75
Leucine synthesis from 4-monoisopentanone	Amino Acid Metabolim	70
Attachment of GlcNAc to undecaprenol	LPS synthesis pathway	67
N-acetylglucosamine-6-phosphate deacetylation	Sugar biosynthesis	61
Succinyl-CoA synthesis	TCA cycle	43
Citrate synthase	TCA cycle	30

Experimental Verification of Knockout Studies

To determine whether our determined knockouts affected protein glycosylation in real cells, we performed an assay for glycan production. It is known that there exists crosstalk between the LPS synthesis pathway and the glycosylation pathway in *E. coli* [67]. after glycan is flipped to the periplasmic side of the inner membrane, that some is transferred to the lipid carrier, lipid A, and exported to the surface of the cell. We reasoned that production of the glycan on the cell surface would be a reasonable measure of total glycan production, which in turn would reasonably approximate protein glycosylation.

We selected four gene knockouts from the Keio strain collection, which contains single, in-frame knockouts of almost every identified gene in *E. coli* [7]

and transformed them with the pACYC-pgl plasmid containing the glycosylation locus from *C. jejuni*. Many of the reactions we identified as either positively or negatively affecting glycosylation are either essential for cell growth or are catalyzed by multiple isozymes, making single-knockout study difficult. The knockouts selected were (1) *acnB* (acotinase B) which catalyzes citrate isomerization as part of the TCA cycle, (2) *gltA* (citrate synthase), which synthesizes citrate from oxaloacetate, (3) *pck* (pyruvate carboxykinase), which is a the first step in gluconeogenesis, and (4) *rpiA*, which isomerizes ribulose-5-phosphate to ribose-5-phosphate for inclusion into the pentose phosphate pathway. We then grew cells to mid-log phase to approximate the steady state that is simulated in the flux balance analysis model.

Cells at mid-log phase were spun down and resuspended in AlexaFluor-conjugated soybean agglutinin (SBA). SBA binds terminal galactose and N-acetylgalactosamine residues, so glycans on the cell surface will be fluorescently labeled for analysis by flow cytometry. We also selected this method because it is easily adaptable to high throughput studies.

The results of this experiment can be found in Figure 4.3. We expected that knockout of acotinase B and citrate synthase would increase glycosylation by decreasing flux away from oxaloacetate. Unexpectedly, we found that the knockouts produced the opposite effect, decreasing fluorescence by roughly 2.5- and 5-fold, respectively. Knockouts of pyruvate carboxykinase and ribulose phosphate isomerase were expected to decrease glycosylation by decreasing gluconeogenesis and flux to the pentose phosphate pathway respectively [39]. In these cases, fluorescence did decrease roughly 5- and 4.5-fold. This result confirmed that our model was able to predict the behavior of some knockouts but

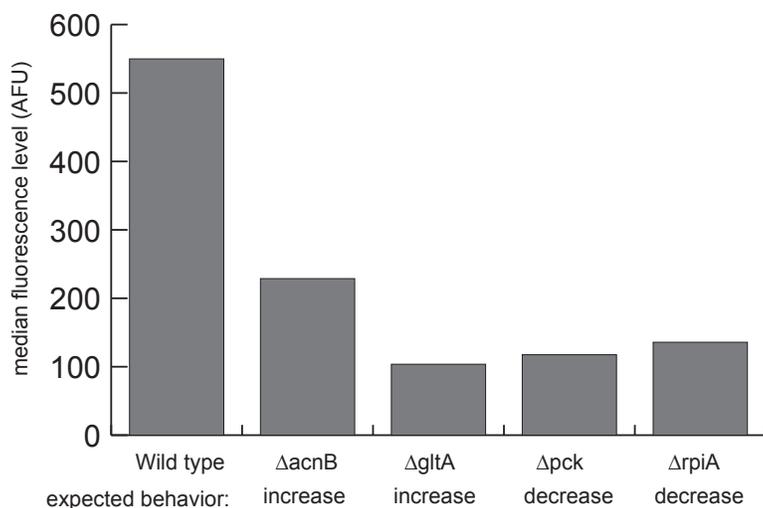


Figure 4.3: Reactions of the central carbon metabolism that are coupled to protein glycosylation. Arrows in bold indicate coupled reactions.

not others.

A key weakness of flux balance analysis is the absence of parameters for metabolic regulation. Often cells with a gene knocked out can compensate by upregulating production of other pathways. The application of regulation would certainly improve the chances of predicting cell behavior during glycoprotein production. Also, the steady-state nature of flux balance analysis does not take into account kinetic parameters, which could be found to be important in the glycosylation process.

We have adapted a genome-scale model of *E. coli* to account for *N*-linked glycosylation using the *C. jejuni* glycosylation pathway. The TCA cycle, glyoxylate cycle, and gluconeogenesis were found to be important in growth and protein

glycosylation in the simulated *E. coli*. Two gene knockouts that we predicted to decrease glycosylation did accomplish this in cells. However, two knockouts that we predicted to increase glycosylation did not result in increased glycosylation compared to wild type. Further study of this network is needed to elucidate factors that can increase the efficiency of glycosylation.

Materials and Methods

Flux Balance Analysis Modeling

The reactions of the iAF1260 model were imported via the Systems Biology Markup Language (SBML). Reactions added were constructed in SBML and added to the iAF1260 model. The SBML was then used to generate a stoichiometric matrix with the UNIVERSAL code generation software. Reactions added consisted of sugar biosynthesis, flipping, and transfer of the glycan to a protein of interest. The adapted model accounted for 3600 reactions, 1290 genes, and 1972 species segregated into 3 compartments: cytoplasm, periplasm, and extra-cellular space. For a list of the reactions added to the model, see Table

The model was optimally solved by linear programming using the glpk solver in OCTAVE. In cases where the maximization of biomass was not the objective function, the model was constrained to produce biomass at 90% of the maximum rate given the growth conditions.

Because enzymes are largely implicit in the flux balance analysis model, a protein species named "targetprotein" was created to act as an acceptor for the glycan. Production of this protein was not factored into the biomass ob-

Table 4.3: Reactions added to the *E. coli* iAF1260 model to study the metabolic effects of glycosylation

Enzyme	Reaction	Reference
Gne	UDP-GlcNAc/UDP-GalNAc epimerase	[16]
PglF	UDP-GlcNAc dehydratase (Bac synthesis)	[200]
PglE	UDP-2-acetamido-2,6-dideoxy- α -d-xylo-4-hexulose amino-transferase (Bac synthesis)	[200]
PglD	N-acetyltransferase (Bac synthesis)	[175]
PglC	Bacillosamine transferase to undecaprenol	[83]
PglA	GalNAc transferase to undecaprenol-Bac	[247]
PglH	GalNAc transferase (3 reactions)	[229]
PglJ	GalNAc transferase	[82]
PglI	Glc transferase	[114]
PglK	Flippase (ATP-driven)	[114]
PglB	Promiscuous oligosyltransferase	[145]
WaaL ¹	Transfer of glycan to Lipid A	[67]
wzx ¹	Promiscuous flippase (Non-ATP-driven)	[1]

jective function. Glucose/minimal media conditions were analyzed as in [66]. Simulated LB media was defined as the minimal media salts from the glucose/minimal media conditions, plus the 20 amino acids as a carbon source.

Knockout and Overexpression Studies

For knockout studies, we first solved the model with no constraints and an objective function of maximization of biomass. We then constrained for the production of biomass and set the objective function to maximization of glycan flipping into the periplasm while constrained each individual reaction flux to zero, one

at a time, simulating knockdown of the gene responsible for this reaction. Reactions that reduced glycan production significantly were noted.

For overexpression studies, we first solved the model with no constraints and an objective function of maximization of biomass. Next, we first solved the model with no constraints and an objective function of maximization of each individual reaction. We then constrained for the production of biomass and set the objective function to maximize each individual flux. With this information, we solved the model again, this time constraining for both 90% biomass production and maximum flux at each reaction. The objective function was the maximum glycosylation of protein with the *C. jejuni* glycan. Reactions that decreased glycosylation when significantly "overexpressed" were noted. The flux coupling finder (FCF) algorithm of Burgard et al. [31] was employed to determine reactions coupled to the production of glycosylated protein.

Analysis of glycan production

Knockout strains from the Keio strain collection [7] were transformed with the pACYC184-pgl plasmid (kindly provided by Markus Aebi) and grown into mid-log phase on LB media supplemented with 0.50% glycerol $\mu\text{g}/\text{mL}$ chloramphenicol. Glycan production was measured by fluorescent labeling of glycan on the cell surface by staining with fluorescent soybean agglutinin (SBA)-AlexaFluor 488 nm conjugate (Invitrogen). Cells were resuspended in 50 $\mu\text{g}/\text{mL}$ SBA-AlexaFluor in PBS, incubated for 1 hour, then washed with 0.5 mL PBS before analysis on a BD FACScalibur flow cytometer.

CHAPTER 5

AN ASSAY FOR REPORTING PROTEIN-PROTEIN INTERACTIONS IN THE BACTERIAL PERIPLASM

Introduction

Protein-protein interactions abound in living cells and understanding in vivo protein interactions is key to understanding many cellular functions. Several tools currently exist to detect protein-protein interactions. The most widely used is the yeast two-hybrid assay, developed in 1989 [69]. Briefly, this assay comprises two proteins fused to a split yeast transcription factor (originally GAL4) which binds a promoter upstream of a reporter protein. If the proteins interact, the activity of the transcription factor is reconstituted and transcription of the reporter protein is upregulated, providing a signal. The yeast two-hybrid assay is highly versatile and is still widely used for analysis of the complex interactions of eukaryotic cellular networks. However, it has its drawbacks, including its location in the nuclear environment of the eukaryotic yeast host, which may differ from the in vivo interaction environment of the proteins of interest.

In recent years, an alternative to the yeast two-hybrid assay has arisen in the form of the protein complementation assay. This method fuses the proteins of interest to a split reporter protein such as GFP [34, 32], YFP [25], luciferase [120], dihydrofolate reductase [191], or β -lactamase [78, 249]. The laboratory of Stephen Michnick (Biochemistry, University of Montreal) has developed, using protein engineering techniques, many split proteins whose individual fragments are inactive. Upon interaction of the proteins of interest the split reporter

protein regains its activity. **Figure 5.1** shows a schematic of the PCA. In 2002, both the Michnick lab [78] and the laboratory of Helen Blau (Stanford) [249] published versions of a split β -lactamase protein complementation assay, both intent on monitoring protein-protein interactions in mammalian cells using Bla. Both constructs were produced in *E. coli*. The Michnick lab construct, containing no signal sequence, measured Bla activity in vitro by nitrocefin colorimetric assay. The Blau lab incorporated Sec signal peptides into both fragments and plated cells on ampicillin. This was the first instance of in vivo protein complementation measured in the periplasm of *E. coli*. Both groups then went on to study protein interactions in mammalian cells using signal-less Bla.

The periplasm has reduced molecular crowding, high sensitivity to extracellular conditions which allows simple modification of the local environment, and an oxidizing milieu that permits disulfide bond formation. In addition, post-translational modifications that are important for producing human-like therapeutic proteins, such as glycosylation, have been demonstrated in the periplasm [239]. Finally, full-length human immunoglobulins (IgGs), which require disulfide bond formation for functional expression, have been successfully expressed in the periplasm of *E. coli* [157]. Cytoplasmic expression of antibody fragments is unfavorable, often requiring several rounds of directed protein evolution to eliminate the need for disulfide bonding. Even in these cases, single-chain variable fragment (scFv) expression and activity is far below that of scFvs expressed in the periplasm.

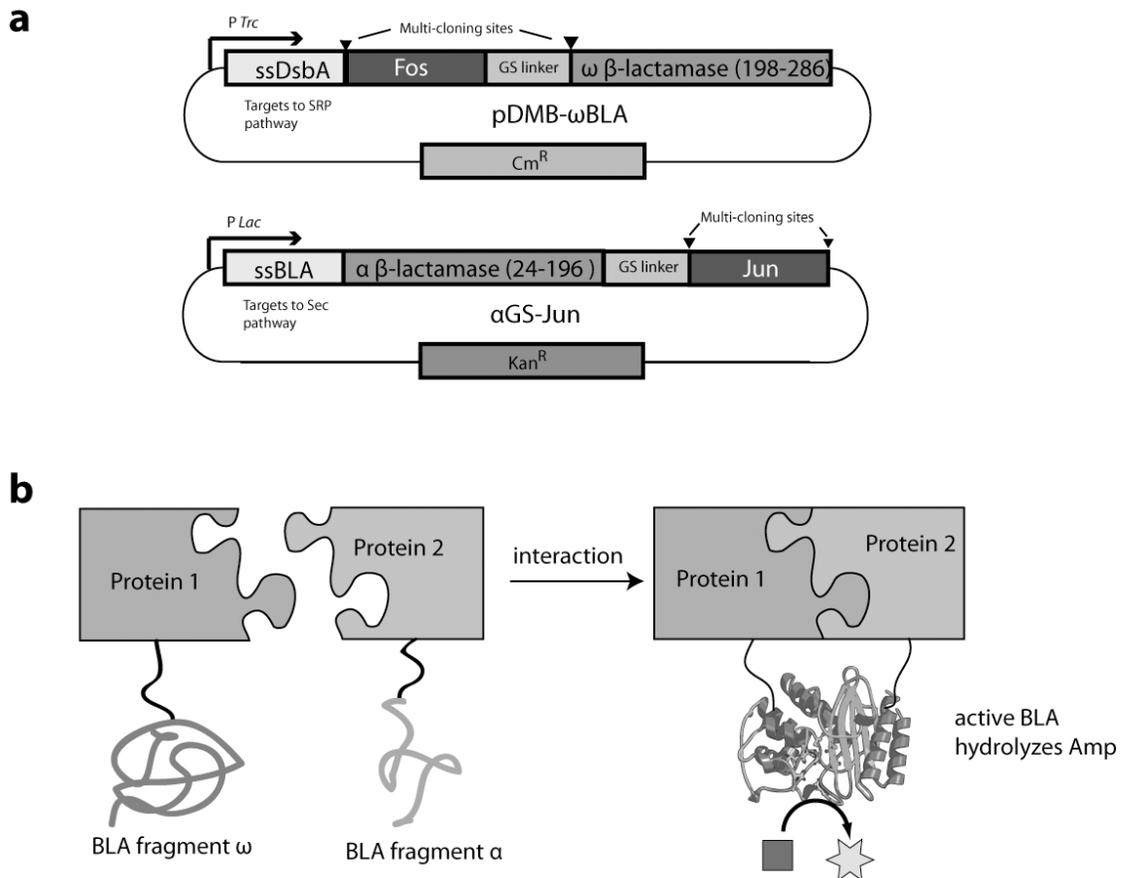


Figure 5.1: (a) Schematic of the plasmids used in this work. The reporter is a two-plasmid system in which genes of interest are fused to the inactive fragments of BLA. Fos and Jun are used here as examples, and proteins of interest would replace Fos and Jun in their respective constructs. (b) Schematic of the protein complementation assay. Protein interactions result in formation of an active BLA complex which is able to hydrolyze ampicillin, conferring an antibiotic-resistance phenotype.

Results and Discussion

We hypothesized that detecting protein interactions in the periplasm can be used to engineer antibody fragments with in vivo activity in the periplasm. The periplasm is ideal for antibody engineering as its oxidizing environment allows for the formation of disulfide bonds, which many antibody fragments require to function [6]. Our approach was to fuse known interacting protein

domains to genes encoding 2 split β -lactamase fragments using a two-plasmid system, the templates for which were kindly given by Helen Blau [249] (see **Figure 5.1a**). The α fragment encodes residues 1-196 of TEM-1 Bla, while the ω fragment encodes residues 198-286 out of 286. The α fragment contains the native signal sequence of Bla (residues 1-23), which targets the C-terminally fused protein to the Sec translocon for transport into the periplasm. In our construct, the protein of interest fused to the ω fragment is targeted to the SRP pathway for co-translational translocation.

Antibody-Antigen Interactions are reported in 2 cases

To determine the ability of our assay to detect and eventually engineer antibody-antigen interactions, we cloned interacting protein pairs as well as non-interacting proteins into two plasmids carrying the two inactive BLA fragments (**Figure 5.1a**). In this study we examined two antibody-antigen pairs. First, we tested an scFv which is known to bind the GCN4 leucine zipper. In addition to the scFv and wild-type GCN4 pair, we also tested a double proline mutant, GCN4-PP as an antigen for the scFv. GCN4 is normally a homodimer, but a double proline mutant of GCN4 (GCN4-pp) exhibits reduced dimerization, thereby reducing the amount of inactive α/α dimers in our assay. **Figure 5.2** shows the results of spot plating of the interacting pairs, with α -Jun as a negative control. The spot plates correspond to a 16-fold higher MBC (minimum bacteriocidal concentration, see Materials and Methods for full definition of MIC/MBC) on ampicillin for GCN4-PP and a 2-fold higher MBC for the GCN4 wild-type (homodimer). A difference of this magnitude is promising for future library selection processes.

In the case of GCN4, we had already tested an antibody fragment that bound a leucine zipper, a short peptide fragment. Since a goal of this assay is engineering protein-protein interactions, we wanted to examine an antibody fragment with a protein antigen. We chose D10, an scFv with affinity for the phage protein gpD, an 11.4 kDa capsid protein from bacteriophage lambda [124]. Indeed, our results with D10 were similar (spot plating shown in **Figure 5.2**), with an MBC roughly 8-fold higher than the Jun negative control and 16-fold higher than the GCN4-PP negative control.

Leucine zipper interactions are reported reliably

The hydrophobic interaction between Fos and Jun leucine zippers is well-documented and these proteins are often used as standards for protein interaction[249]. The K_D of their interaction has been reported in the nanomolar range [183]. In our study, we fused Fos to the ω fragment and Jun to the α fragment. As a negative control, we incorporated as a potential binding partner the leucine zipper GCN4, which does not bind Fos. When placed in the PCA, Fos and Jun were found to have ampicillin activity well above that of the negative control **Figure 5.2b**. After successful spot plating, we further characterized the interactions by constructing a series of point mutants which have been shown to have lower affinity (due to knockout of leucine residues). The MIC/MBC of these interactions and spot plating of these mutants can be found in **Figure 5.3**. It is notable that our assay is apparently able to correlate Bla activity with levels of interaction along a spectrum of affinities. Further study of the expression and actual affinities of these variants is warranted To verify that the intermediate Bla activity is indeed due to intermediate affinities and not changes in solubility of

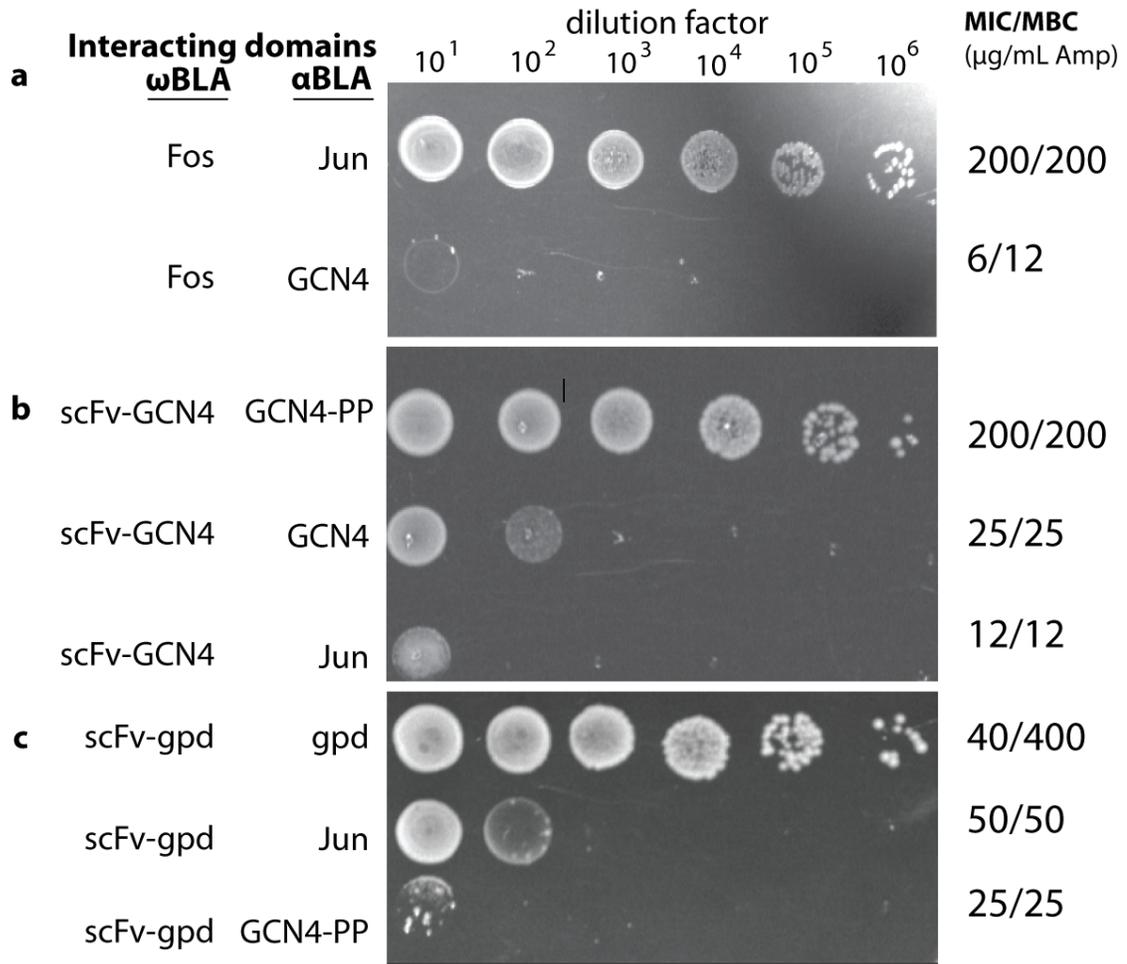


Figure 5.2: **Protein complementation of interacting domains.** (a) Spot plating of Fos and Jun leucine zippers on $50 \mu\text{g/ml Amp}$. GCN4, another leucine zipper, but one that does not interact with Fos, is used as a negative control. (b) Spot plating as in (a) of antibody fragment scFv-GCN4 and its antigen, the GCN4 leucine zipper. GCN4-PP is a double point mutant to wild-type GCN4 which reduces homodimerization of the leucine zippers. Jun is used as a negative control. (c) Spot plating as in (a) of scFv D10[124], with affinity to phage protein Gpd. Jun and GCN4 used as negative controls.

the point mutants.

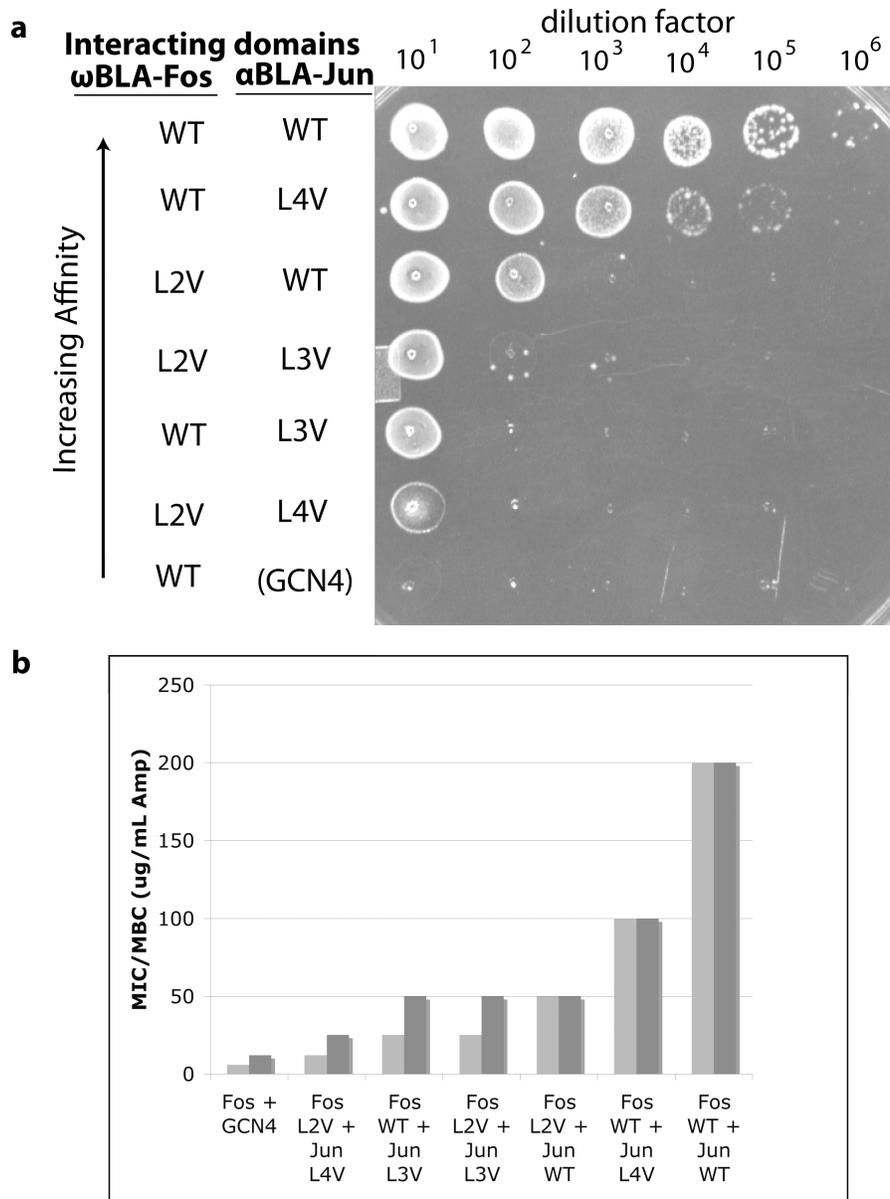


Figure 5.3: **Interaction of Fos and Jun variants** (a) Spot plating as in Figure 5.2 of Fos/Jun variants. Note: notations such as L3V refer replacement of leucine 3 with valine, not necessarily residue 3. Similar growth for all variants was confirmed on Cm/Kan plates (data not shown) (b) MIC (blue) and MBC (red) measurements of Fos and Jun variants pictured in (a).

Continued development of this assay is warranted to optimize plating Amp

concentration, number of cells, etc. for the inclusion of selections for engineered protein interactions. The considerable difference in MIC/MBC between the GCN4 and Jun cases with scFv(GCN4) implies that selection of interactions from a randomized or even naïve group of proteins is not far away.

Materials and Methods

Construction of vectors

All cloning was performed using standard molecular biological techniques. A gene sequence comprising a (GGGGS)₃-NGR linker sequence followed by residues 198-286 of TEM-1 BLA was cloned between the BamHI and HindIII sites of pDMB (Mansell et al., manuscript in preparation) to create pDMB- ω BLA. Genes to be fused to the ω BLA fragment were then cloned into this vector between XbaI and Sall sites. Genes to be fused to the α BLA fragment were cloned between the KpnI and BamHI sites of vector α GS-Jun [249]. Fos and Jun leucine knockouts and scFv-GCN4 were as in [245]. Template DNA for scFvs D10, and *gpd* was kindly provided by Andreas Plückthun.

Plasmid pMAZ360 was kindly provided by George Georgiou. To create α -IgG(GCN4), we first digested pMAZ360-YMF10 with DraI to remove the *bla* gene from this vector. Chloramphenicol acetyltransferase (*cat*) was ligated in its place by blunt-end ligation to create pMAZ*cat*. Next, we designed DNA (synthesized by Genscript) that would replace the variable regions VL and VH of the YMF10 IgG with α BLA-VL(GCN4) and VH(GCN4). We ligated this DNA fragment into pMAZ*cat* with XbaI and HindIII to create α -IgG(GCN4). Plas-

mid $p\omega$ -GCN4-PP was created by ligation of the GCN4- ω BLA fragment from pDMB- ω BLA between the SphI and ClaI sites in pLYNK-YFP1, a spectinomycin-resistant vector described elsewhere [126].

Expression of fusion proteins and cell growth assays

E. coli MC4100 cells were co-transformed with plasmids pDMB-P1- ω BLA (CmR) and aGS-P2 (KanR), where P1 and P2 represent proteins of interest. Cells were grown overnight at 37C in LB medium containing 20 μ g/mL chloramphenicol (Cm) and 50 μ g/mL kanamycin (Kan) or 50 μ g/mL spectinomycin. Screening of cells on LB agar was performed by first normalizing overnight cultures by OD600 and then spotting 5 μ L of serially diluted (10-10-fold) cells on LB agar plates with 12, 25, 50, or 50 μ g/mL of ampicillin (Amp). In all cases, the plates were incubated 16 h at 37C and then imaged using a ChemiDoc System (Bio-Rad). For MBC/MIC determination, approximately 200 colony forming units (CFUs) of each clone were plated on LB agar plates containing 0, 3, 6, 12, 25, 50, 100, 200, 400, 800, or 1600 μ g/mL Amp or 20 μ g/mL Cm. The minimum bacteriocidal concentration was determined as the minimum Amp concentration at which no colonies appeared on the plates. Minimum inhibitory concentration (MIC) was determined as the minimum concentration of Amp on which colony size was significantly smaller than control.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

Abstract

I have described several methodologies useful for detecting protein solubility, interactions, and glycosylation in the periplasm. The applications of these methods to protein engineering have been demonstrated in some cases (e.g., the engineering of more soluble Amyloid β peptides explained in Chapter 2). This chapter will expand on these tools and propose experiments that could lead to further applications for protein and glycosylation engineering in *E. coli*.

Glycosylation of a eukaryotic protein

We sought to determine if a eukaryotic glycoprotein, RNase, would be glycosylated in the context of a C-terminal TEM-1 β -lactamase fusion, as in Chapter 3. We cloned the gene encoding bovine RNase A, the aglycosylated form, into the pDMB reporter system. A point mutation had been previously introduced into this gene to incorporate an S32D mutation[128]. This mutation allowed the amino acid sequence to correspond to the known *C. jejuni* consensus glycosylation sequence of D/E-X₁-N-X₂-S/T [129].

RNase A has been shown in literature to be more stable when glycosylated. Arnold *et al.* showed that glycosylated RNase A was more resistant to protease degradation, had significantly slower rate constants, and had a 1.5°C higher melting temperature compared to the aglycosylated form [5]. Molecular dy-

namics studies of the protein have shown that glycosylation destabilizes the unfolded state via the formation of hydrogen bonds during the unfolding process. The same study also predicted that glycosylation adds 2.5 kJ/mol to the overall protein stability [42]. We reasoned that this increase in stability due to glycosylation could translate to increased growth on ampicillin in the context of a BLA fusion, as we had previously seen in the case of Peb3 (Chapter 3).

We co-transformed the plasmid expressing the RNase-BLA fusion with either pglKan or pglmutKan (see Chapter 3 for descriptions). Spot plating on ampicillin revealed no change in overall growth phenotype due to glycosylation (Figure 6.1a). A nitrocefin hydrolysis assay for BLA activity also showed no significant difference (Figure 6.1c). We hypothesized that that glycosylation of this eukaryotic protein was inefficient. Indeed, the glycosylation site of RNase lies in a highly structured region [128] which is difficult to glycosylate using the enzymes of the *C. jejuni* pathway.

Western blotting confirmed that RNase is glycosylated in the context of an RNase-BLA fusion (Figure 6.1b). To our knowledge, this is the first instance of soluble production of glycosylated RNase A in *E. coli*. However, the efficiency of glycosylation was not apparent from the blot. Further experiments are needed to determine the actual glycosylation efficiency. Nitrocefin hydrolysis assay for BLA activity also showed no significant difference (Figure 6.1c).

We next sought to determine whether cotranslational translocation contributed to our ability to glycosylate this protein, as the DsbA signal sequence that is N-terminal to the fusion has been shown to co-translationally translocate proteins [199]. To test this hypothesis, we made fusions of the signal sequences of PhoA, MalE, and DsbA to RNaseA. The ssPhoA and ssMalE are directed

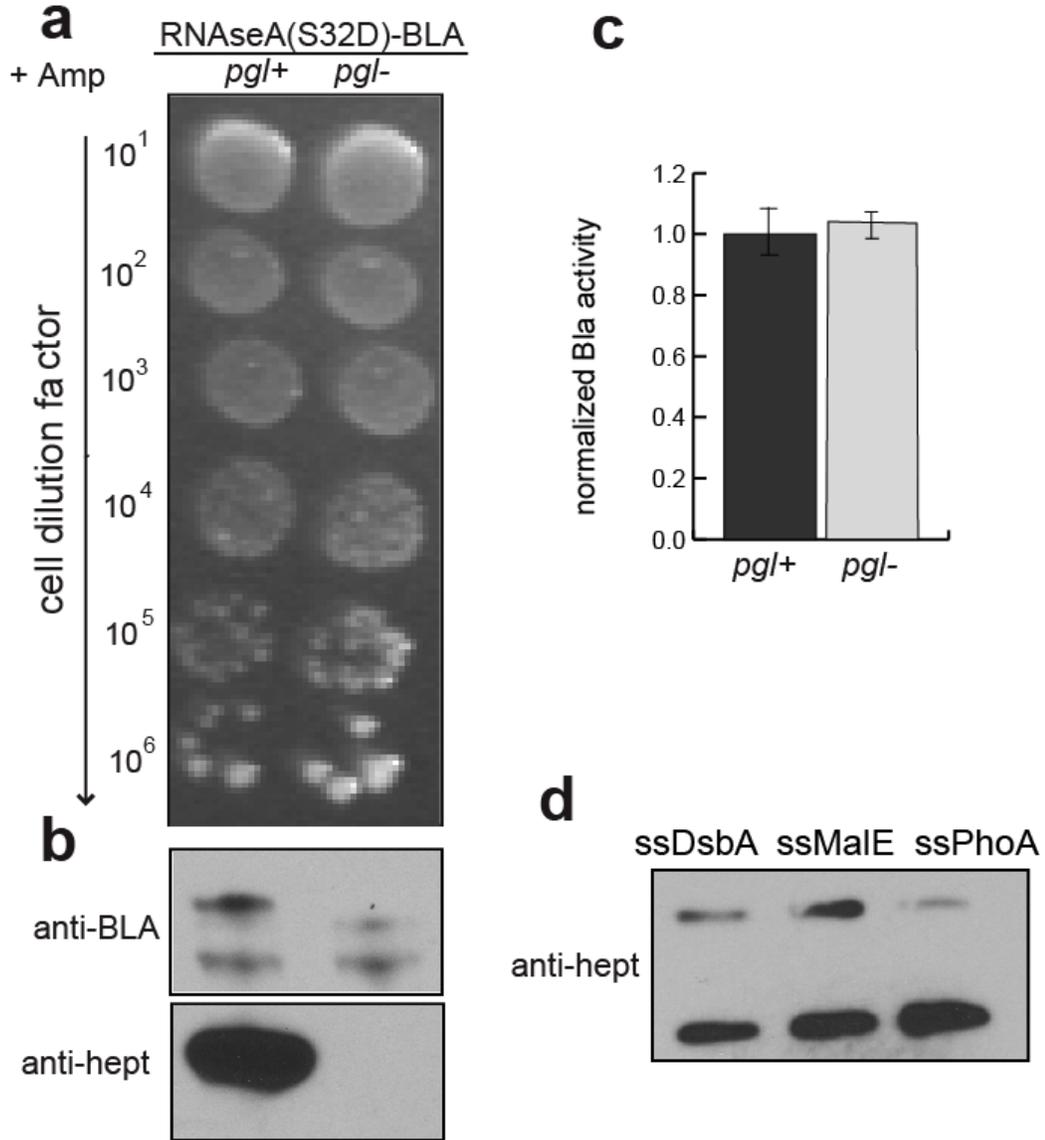


Figure 6.1: Analysis of RNase A-BLA fusions. (a) Spot plating as in Figure 3 on 200 $\mu\text{g}/\text{mL}$ Amp and 50 $\mu\text{g}/\text{mL}$ Kan (b) Western blot analysis as in Figure 3 (c) Nitrocefin hydrolysis activity as in Figure 3. (d) Western blotting with hR6 of RNaseA (no BLA fusion) fused to N-terminal signal sequences of DsbA, MalE, and PhoA.

through the Sec pathway, with ssMaleE export being dependent on SecB [135] and ssPhoA export independent of SecB [134]. Western blotting of the signal sequence-RNaseA S32D fusions showed that glycosylation occurred in all three cases. This implies that co-translational translocation may not affect the overall efficiency of glycosylation in the case of this particular protein.

Materials and Methods

The gene for RNaseA S32D was amplified from the plasmid pMIK81, generously provided by Markus Aebi [128]. For solubility studies, RNase was cloned between the SfiI sites of pDMB (as described in Chapter 3). For signal sequence studies, plasmids based on pTrc99A containing the ssPhoA, ssMaleE, and ssDsbA signal sequences were used as backbone [72] and RNaseA was cloned between XbaI and HindIII sites in these plasmids. A FLAG tag was appended to RNase A 3' of the HindIII site and stop codon for immunodetection. Spot plating, Western blotting, and nitrocefin assays were performed as described in Chapter 3.

Engineering glycosylation in *E. coli*

The methodology described in Chapter 3, a reporter of protein glycosylation, has obvious applications for directed evolution and glycoengineering. Using the ampicillin resistance phenotype as a reporter, one could imagine performing selections (1) for functional glycosylation machinery and (2) for improved glycosylation machinery. For example, the method could be used to study the efficiency of glycosylation of multiple oligosyltransferases (OSTs) by sim-

ply cloning families of monomeric OSTs in place of the *C. jejuni* PglB. The high throughput of this method would allow preliminary characterization of many OSTs in a single round of plating on ampicillin.

Next, directed evolution via colony PCR, DNA shuffling, or any other comparable method of library creation could be undertaken on the *C. jejuni* *pgl* system to improve the overall rate of glycosylation. In this system, protein glycosylation is likely not at its peak efficiency and as such improved glycosylation in proteins whose solubility is sensitive to sugar attachment (e.g., Peb3, CjaA, and engineered im7 peptides) could lead to increased growth on ampicillin. Those sets of engineered machinery proteins which lead to higher growth on ampicillin could then be further characterized.

Finally, as mentioned in Chapter 4, pathways inherent to *E. coli* metabolism could be manipulated to affect glycosylation efficiency. By screening knockout strains (or libraries of knockouts such as the Keio collection [7]) or libraries created by such strain engineering methods as MAGE [242] or TRMR [246], it could be possible to isolate genetic elements of bacterial cells that contribute to glycosylation and use this information to optimize the overall production of glycoproteins.

A reporter of protein-carbohydrate interactions

To our knowledge, there exists no method of reporting the interactions between proteins and carbohydrates *in vivo*. Combining some of the tools developed in this work could lead to the creation of such a reporter. A method of reporting specific protein-carbohydrate interactions would be useful in many ways. First,

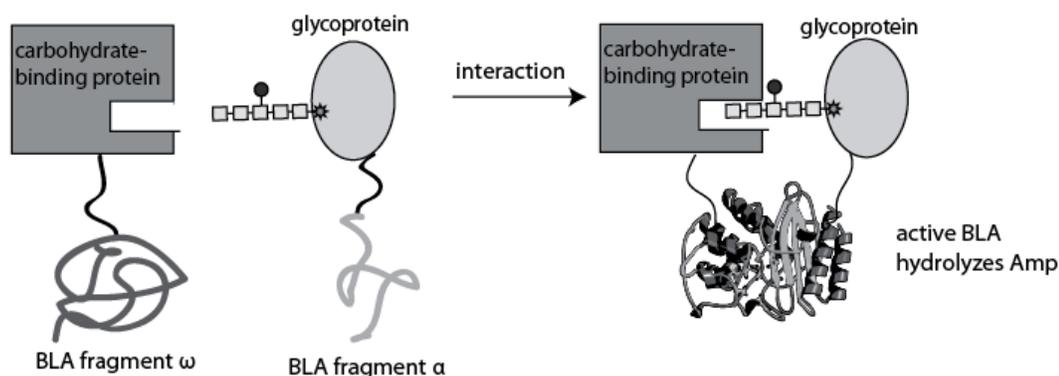


Figure 6.2: Schematic of the protein-carbohydrate interaction complementation assay. Binding of a sugar group by a carbohydrate binding protein results in formation of an active BLA complex which is able to hydrolyze ampicillin, conferring an antibiotic-resistance phenotype.

it could provide rapid, direct evidence of glycosylation as opposed to the indirect method of glycosylation reporting (via a change in folding and solubility) detailed in Chapter 3. Second, by designing a reporter to interact with specific glycans, the reporter could be used for engineering of these glycans. Third, the method could be used to evolve novel or more specific protein-carbohydrate interactions.

Figure 6.2 shows a schematic of how such a protein-carbohydrate reporter might operate in *E. coli*. A carbohydrate binding domain (CBD) (e.g., a lectin or glycan-specific antibody) is fused to an inactive split fragment of TEM-1 β -lactamase, much in the same manner as described in Chapter 5. The second split fragment of BLA is fused to a target glycoprotein. When this target protein is glycosylated with sugars to which the CBD has some affinity, the binding of the CBD and glycan would complement BLA activity. Selection for this binding activity can be accomplished by plating on ampicillin or other β -lactam antibiotics.

Conclusions

We have presented tools that will help us to better understand and engineer the environment of the *E. coli* periplasm. Protein folding, solubility, interactions, and post-translational modifications in this compartment are required for the economical production of many therapeutic proteins in *E. coli*. Specifically, the methods presented in this work have laid the groundwork for engineering of glycoprotein production in *E. coli*. Further experiments (explained above) could lead to specific engineering of cellular glycan structure and improvement of carbohydrate-binding domains. In the end, gaining understanding of the complex processes of protein translocation, folding, and post-translational modifications in the periplasm is critical to solving an array of biotechnological problems and further expands the role of the well-understood workhorse organism *Escherichia coli*.

BIBLIOGRAPHY

- [1] Cristina Alaimo, Ina Catrein, Laura Morf, Cristina L Marolda, Nico Callewaert, Miguel A Valvano, Mario F Feldman, and Markus Aebi. Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J*, 25(5):967–976, 2006.
- [2] Hal Alper, Curt Fischer, Elke Nevoigt, and Gregory Stephanopoulos. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A*, 102(36):12678–12683, 2005.
- [3] J. P. Arie, M. Miot, N. Sassoon, and J. M. Betton. Formation of active inclusion bodies in the periplasm of escherichia coli. *Mol Microbiol*, 62(2):427–37, 2006. 0950-382X (Print) Journal Article.
- [4] J. P. Arie, N. Sassoon, and J. M. Betton. Chaperone function of fkpa, a heat shock prolyl isomerase, in the periplasm of escherichia coli. *Mol Microbiol*, 39(1):199–210, 2001. 0950-382X (Print) Journal Article Research Support, Non-U.S. Gov't.
- [5] U Arnold, A Schierhorn, and R Ulbrich-hofmann. Modification of the unfolding region in bovine pancreatic ribonuclease and its influence on the thermal stability and proteolytic fragmentation. *Eur J Biochem*, 259(1-2):470–475, 1999.
- [6] A. Auf der Maur, D. Escher, and A. Barberis. Antigen-independent selection of stable intracellular single-chain antibodies. *FEBS Lett*, 508(3):407–12, 2001.
- [7] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of escherichia coli k-12 in-frame, single-gene knockout mutants: the keio collection. *Mol Syst Biol*, 2:2006.0008, 2006.
- [8] A. M. Baca and W. G. Hol. Overcoming codon bias: a method for high-level overexpression of plasmodium and other at-rich parasite genes in escherichia coli. *Int J Parasitol*, 30(2):113–8, 2000.
- [9] Tania A Baker and Robert T Sauer. Atp-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci*, 31(12):647–653, 2006.

- [10] F. Baneyx. Recombinant protein expression in escherichia coli. *Curr Opin Biotechnol*, 10(5):411–21, 1999. 0958-1669 (Print) Journal Article Research Support, U.S. Gov't, Non-P.H.S. Review.
- [11] F. Baneyx and M. Mujacic. Recombinant protein folding and misfolding in escherichia coli. *Nat Biotechnol*, 22(11):1399–408, 2004. 1087-0156 (Print) Journal Article Research Support, U.S. Gov't, Non-P.H.S. Review.
- [12] D. Baram, E. Pyetan, A. Sittner, T. Auerbach-Nevo, A. Bashan, and A. Yonath. Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proc Natl Acad Sci U S A*, 102(34):12017–22, 2005.
- [13] J. Beckwith. What lies beyond uranus? preconceptions, ignorance, serendipity and suppressors in the search for biology's secrets. *Genetics*, 176(2):733–40, 2007. 0016-6731 (Print) Journal Article.
- [14] I. Benz and M.A. Schmidt. Never say never again: protein glycosylation in pathogenic bacteria. *Molecular Microbiology*, 45(2):267–276, 2002.
- [15] H. Berges, E. Joseph-Liauzun, and O. Fayet. Combined effects of the signal sequence and the major chaperone proteins on the export of human cytokines in escherichia coli. *Appl Environ Microbiol*, 62(1):55–60, 1996.
- [16] S. Bernatchez, C.M. Szymanski, N. Ishiyama, J. Li, H.C. Jarrell, P.C. Lau, A.M. Berghuis, N.M. Young, and W.W. Wakarchuk. A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. *Journal of Biological Chemistry*, 280(6):4792, 2005.
- [17] P. H. Bessette, F. Aslund, J. Beckwith, and G. Georgiou. Efficient folding of proteins with multiple disulfide bonds in the escherichia coli cytoplasm. *Proc Natl Acad Sci U S A*, 96(24):13703–8, 1999. 0027-8424 (Print) Journal Article Research Support, U.S. Gov't, P.H.S.
- [18] J. M. Betton and M. Hofnung. Folding of a mutant maltose-binding protein of escherichia coli which forms inclusion bodies. *J Biol Chem*, 271(14):8046–52, 1996. Betton, J M Hofnung, M United states The Journal of biological chemistry J Biol Chem. 1996 Apr 5;271(14):8046-52.
- [19] E. T. Boder and K. D. Wittrup. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol*, 15(6):553–7, 1997.

- [20] F. Bonekamp and K. F. Jensen. The agg codon is translated slowly in e. coli even at very low expression levels. *Nucleic Acids Res*, 16(7):3013–24, 1988.
- [21] Carlos J Bosques and Barbara Imperiali. The interplay of glycosylation and disulfide formation influences fibrillization in a prion protein fragment. *Proc Natl Acad Sci U S A*, 100(13):7593–7598, 2003.
- [22] H. Bothmann and A. Pluckthun. Selection for a periplasmic factor improving phage display and functional periplasmic expression. *Nat Biotechnol*, 16(4):376–80, 1998. 1087-0156 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [23] H. Bothmann and A. Pluckthun. The periplasmic escherichia coli peptidylprolyl cis,trans-isomerase fkpa. i. increased functional expression of antibody fragments with and without cis-prolines. *J Biol Chem*, 275(22):17100–5, 2000.
- [24] G A Bowden, A M Paredes, and G Georgiou. Structure and morphology of protein inclusion bodies in escherichia coli. *Biotechnology (N Y)*, 9(8):725–730, 1991.
- [25] Keren Bracha-Drori, Keren Shichrur, Aviva Katz, Moran Oliva, Ruthie Angelovici, Shaul Yalovsky, and Nir Ohad. Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. *Plant J*, 40(3):419–427, 2004.
- [26] Pascal Braun, Yanhui Hu, Binghua Shen, Allison Halleck, Malvika Koundinya, Ed Harlow, and Joshua LaBaer. Proteome-scale purification of human proteins from bacteria. *Proc Natl Acad Sci U S A*, 99(5):2654–2659, 2002.
- [27] J K Broome-Smith, L D Bowler, and B G Spratt. A simple method for maximizing the yields of membrane and exported proteins expressed in escherichia coli. *Mol Microbiol*, 3(12):1813–1817, 1989.
- [28] P. N. Bryan. Protein engineering of subtilisin. *Biochim Biophys Acta*, 1543(2):203–222, 2000.
- [29] J. E. Bryant, J. T. Lecomte, A. L. Lee, G. B. Young, and G. J. Pielak. Protein dynamics in living cells. *Biochemistry*, 44(26):9275–9, 2005. 0006-2960 (Print) Journal Article Research Support, N.I.H., Extramural Research

Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S.

- [30] B. Bukau and A. L. Horwich. The hsp70 and hsp60 chaperone machines. *Cell*, 92(3):351–66, 1998.
- [31] Anthony P Burgard, Evgeni V Nikolaev, Christophe H Schilling, and Costas D Maranas. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res*, 14(2):301–312, 2004.
- [32] S. Cabantous, T. C. Terwilliger, and G. S. Waldo. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol*, 23(1):102–7, 2005.
- [33] S. Cabantous and G. S. Waldo. In vivo and in vitro protein solubility assays using split gfp. *Nat Methods*, 3(10):845–54, 2006.
- [34] Stephanie Cabantous, Jean-Denis Pedelacq, Brian L Mark, Cleo Naranjo, Thomas C Terwilliger, and Geoffrey S Waldo. Recent advances in gfp folding reporter and split-gfp solubility reporter technologies. application to improving the folding and solubility of recalcitrant proteins from mycobacterium tuberculosis. *J Struct Funct Genomics*, 6(2-3):113–119, 2005.
- [35] J. H. Cate, M. M. Yusupov, G. Z. Yusupova, T. N. Earnest, and H. F. Noller. X-ray crystal structures of 70s ribosome functional complexes. *Science*, 285(5436):2095–104, 1999.
- [36] Eda Celik, Adam C Fisher, Cassandra Guarino, Thomas J Mansell, and Matthew P DeLisa. A filamentous phage display system for n-linked glycoproteins. *Protein Sci*, 19(10):2006–2013, 2010.
- [37] J. J. Chalmers, E. Kim, J. N. Telford, E. Y. Wong, W. C. Tacon, M. L. Shuler, and D. B. Wilson. Effects of temperature on escherichia coli overproducing beta-lactamase or human epidermal growth factor. *Appl Environ Microbiol*, 56(1):104–111, 1990.
- [38] C. Chamberlain and K. M. Hahn. Watching proteins in the wild: fluorescence methods to study protein dynamics in living cells. *Traffic*, 1(10):755–62, 2000.
- [39] Yun-Peng Chao, Jong-Tzer Chern, Wei Shing Lin, and Zei Wen Wang. Development of a fed-batch fermentation process to overproduce phos-

- phenolpyruvate carboxykinase using an expression vector with promoter and plasmid copy number controllable by heat. *Biotechnol Bioeng*, 84(4):459–466, 2003.
- [40] Helene Chautard, Emilio Blas-Galindo, Thierry Menguy, Laure Grand'Moursel, Felipe Cava, Jose Berenguer, and Marc Delcourt. An activity-independent selection system of thermostable protein variants. *Nat Methods*, 4(11):919–921, 2007.
- [41] Mark M Chen, Alice I Bartlett, Paul S Nerenberg, Claire T Friel, Christian P R Hackenberger, Collin M Stultz, Sheena E Radford, and Barbara Imperiali. Perturbing the folding energy landscape of the bacterial immunity protein im7 by site-specific n-linked glycosylation. *Proc Natl Acad Sci U S A*, 107(52):22528–22533, 2010.
- [42] Youngjin Choi, Jong Hyun Lee, Suntae Hwang, Joong-Kwon Kim, Karpjoo Jeong, and Seunho Jung. Retardation of the unfolding process by single n-glycosylation of ribonuclease a based on molecular dynamics simulations. *Biopolymers*, 89(2):114–123, 2008.
- [43] Bevan Ks Chung, Suresh Selvarasu, Camattari Andrea, Jimyoung Ryu, Hyeokweon Lee, Jungoh Ahn, Hongweon Lee, and Dong-Yup Lee. Genome-scale metabolic reconstruction and in silico analysis of methylo-trophic yeast pichia pastoris for strain improvement. *Microb Cell Fact*, 9:50, 2010.
- [44] L. M. Contreras Martinez and M. P. DeLisa. Intracellular ribosome display via secm translation arrest as a selection for antibodies with enhanced cytosolic stability. *J Mol Biol*, page submitted, 2007.
- [45] L. M. Contreras Martinez, F. J. Martinez-Veracoechea, P. Pohkarel, A. D. Stroock, F. A. Escobedo, and M. P. DeLisa. Protein translocation through a tunnel induces changes in folding kinetics: a lattice model study. *Biotechnol Bioeng*, 94(1):105–17, 2006.
- [46] P. Cortazzo, C. Cervenansky, M. Mari'n, C. Reiss, R. Ehrlich, and A. Deana. Silent mutations affect in vivo protein folding in escherichia coli. *Biochem Biophys Res Commun*, 293(1):537–541, 2002.
- [47] F. Crick. On protein synthesis. *Symp Soc Exp Biol XII*, pages 139–163, 1958.
- [48] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–3, 1970.

- [49] Lindsay M Davis, Tsutomu Kakuda, and Victor J DiRita. A campylobacter jejuni znua orthologue is essential for growth in low-zinc environments and chick colonization. *J Bacteriol*, 191(5):1631–1640, 2009.
- [50] Jr. Del Tito, B. J., J. M. Ward, J. Hodgson, C. J. Gershater, H. Edwards, L. A. Wysocki, F. A. Watson, G. Sathe, and J. F. Kane. Effects of a minor isoleucyl trna on heterologous protein translation in escherichia coli. *J Bacteriol*, 177(24):7086–91, 1995.
- [51] M. P. DeLisa, P. Lee, T. Palmer, and G. Georgiou. Phage shock protein pspa of escherichia coli relieves saturation of protein export via the tat pathway. *J Bacteriol*, 186(2):366–73, 2004. 0021-9193 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [52] M. P. DeLisa, P. Samuelson, T. Palmer, and G. Georgiou. Genetic analysis of the twin arginine translocator secretion pathway in bacteria. *J Biol Chem*, 277(33):29825–31, 2002. 0021-9258 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [53] M. P. DeLisa, D. Tullman, and G. Georgiou. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc Natl Acad Sci U S A*, 100(10):6115–20, 2003. 0027-8424 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S.
- [54] A. I. Derman, W. A. Prinz, D. Belin, and J. Beckwith. Mutations that allow disulfide bond formation in the cytoplasm of escherichia coli. *Science*, 262(5140):1744–7, 1993. 0036-8075 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [55] Clemens Durr, Harald Nothaft, Christian Lizak, Rudi Glockshuber, and Markus Aebi. The escherichia coli glycopage display system. *Glycobiology*, 20(11):1366–1372, 2010.
- [56] M. R. Dyson, S. P. Shadbolt, K. J. Vincent, R. L. Perera, and J. McCafferty. Production of soluble mammalian proteins in escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol*, 4:32, 2004.
- [57] A Edelman, L Bowler, J K Broome-Smith, and B G Spratt. Use of a beta-lactamase fusion vector to investigate the organization of penicillin-binding protein 1b in the cytoplasmic membrane of escherichia coli. *Mol Microbiol*, 1(1):101–106, 1987.

- [58] R. J. Ellis. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol*, 11(1), 2001.
- [59] R. J. Ellis and F. U. Hartl. Principles of protein folding in the cellular environment. *Curr Opin Struct Biol*, 9(1):102–10, 1999.
- [60] Olof Emanuelsson, Soren Brunak, Gunnar von Heijne, and Henrik Nielsen. Locating proteins in the cell using targetp, signalp and related tools. *Nat Protoc*, 2(4):953–971, 2007.
- [61] S. D. Emr, S. Hanley-Way, and T. J. Silhavy. Suppressor mutations that restore export of a protein with a defective signal sequence. *Cell*, 23(1):79–88, 1981.
- [62] A. Erbse, D. A. Dougan, and B. Bukau. A folding machine for many but a master of none. *Nat Struct Biol*, 10(2):84–6, 2003.
- [63] M. S. Evans, K. G. Ugrinov, M. A. Frese, and P. L. Clark. Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat Methods*, 2(10):757–62, 2005.
- [64] O. Fayet, T. Ziegelhoffer, and C. Georgopoulos. The groes and groel heat shock gene products of escherichia coli are essential for bacterial growth at all temperatures. *J Bacteriol*, 171(3):1379–85, 1989.
- [65] B. J. Feilmeier, G. Iseminger, D. Schroeder, H. Webber, and G. J. Phillips. Green fluorescent protein functions as a reporter for protein localization in escherichia coli. *J Bacteriol*, 182(14):4068–76, 2000. 0021-9193 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [66] Adam M Feist, Christopher S Henry, Jennifer L Reed, Markus Krumnacker, Andrew R Joyce, Peter D Karp, Linda J Broadbelt, Vassily Hatzimanikatis, and Bernhard O Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007.
- [67] Mario F Feldman, Michael Wacker, Marcela Hernandez, Paul G Hitchen, Cristina L Marolda, Michael Kowarik, Howard R Morris, Anne Dell, Miguel A Valvano, and Markus Aebi. Engineering n-linked protein glycosylation with diverse o antigen lipopolysaccharide structures in escherichia coli. *Proc Natl Acad Sci U S A*, 102(8):3016–3021, 2005.

- [68] L. Ferbitz, T. Maier, H. Patzelt, B. Bukau, E. Deuerling, and N. Ban. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, 431(7008):590–6, 2004.
- [69] S Fields and O Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246, 1989.
- [70] A. C. Fisher and M. P. DeLisa. A little help from my friends: quality control of presecretory proteins in bacteria. *J Bacteriol*, 186(22):7467–73, 2004. 0021-9193 (Print) Journal Article Research Support, Non-U.S. Gov't Review.
- [71] A. C. Fisher, W. Kim, and M. P. DeLisa. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. *Protein Sci*, 15(3):449–58, 2006.
- [72] Adam C Fisher and Matthew P DeLisa. Laboratory evolution of fast-folding green fluorescent protein using secretory pathway quality control. *PLoS One*, 3(6):e2351, 2008.
- [73] Adam C Fisher and Matthew P DeLisa. Efficient isolation of soluble intracellular single-chain antibodies using the twin-arginine translocation machinery. *J Mol Biol*, 385(1):299–311, 2009.
- [74] Adam C Fisher, Charles H Haitjema, Cassandra Guarino, Eda Celik, Christine E Endicott, Craig A Reading, Judith H Merritt, A Celeste Ptak, Sheng Zhang, and Matthew P DeLisa. Production of secretory and extracellular n-linked glycoproteins in escherichia coli. *Appl Environ Microbiol*, 77(3):871–881, 2011.
- [75] A. M. Flower, R. C. Doebele, and T. J. Silhavy. Prla and prlg suppressors reduce the requirement for signal sequence recognition. *J Bacteriol*, 176(18):5607–14, 1994.
- [76] Stephen S Fong and Bernhard O Palsson. Metabolic gene-deletion strains of escherichia coli evolve to computationally predicted growth phenotypes. *Nat Genet*, 36(10):1056–1058, 2004.
- [77] C. Frech, M. Wunderlich, R. Glockshuber, and F. X. Schmid. Competition between dsba-mediated oxidation and conformational folding of rtm1 beta-lactamase. *Biochemistry*, 35(35):11386–95, 1996. Frech, C Wunderlich, M Glockshuber, R Schmid, F X Research Support, Non-U.S. Gov't United states Biochemistry Biochemistry. 1996 Sep 3;35(35):11386-95.

- [78] A. Galarneau, M. Primeau, L. E. Trudeau, and S. W. Michnick. Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat Biotechnol*, 20(6):619–22, 2002. Galarneau, Andre Primeau, Martin Trudeau, Louis-Eric Michnick, Stephen W Research Support, Non-U.S. Gov't Validation Studies United States Nature biotechnology *Nat Biotechnol*. 2002 Jun;20(6):619-22.
- [79] W. Gao, S. Tyagi, F. R. Kramer, and E. Goldman. Messenger rna release from ribosomes during 5'-translational blockage by consecutive low-usage arginine but not leucine codons in escherichia coli. *Mol Microbiol*, 25(4):707–16, 1997.
- [80] G. Georgiou and L. Segatori. Preparative expression of secreted proteins in bacteria: status report and future prospects. *Curr Opin Biotechnol*, 16(5):538–45, 2005. 0958-1669 (Print) Journal Article Research Support, N.I.H., Extramural Review.
- [81] R. T. Gill, M. P. DeLisa, J. J. Valdes, and W. E. Bentley. Genomic analysis of high-cell-density recombinant escherichia coli fermentation and "cell conditioning" for improved recombinant protein yield. *Biotechnol Bioeng*, 72(1):85–95, 2001.
- [82] Kerney Jebrell Glover, Eranthie Weerapana, and Barbara Imperiali. In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic n-linked glycosylation. *Proc Natl Acad Sci U S A*, 102(40):14255–14259, 2005.
- [83] K.J. Glover, E. Weerapana, M.M. Chen, and B. Imperiali. Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry*, 45(16):5343–5350, 2006.
- [84] E. Goldman, A. H. Rosenberg, G. Zubay, and F. W. Studier. Consecutive low-usage leucine codons block translation only when near the 5' end of a message in escherichia coli. *J Mol Biol*, 245(5):467–73, 1995.
- [85] S. Gottesman. Proteases and their targets in escherichia coli. *Annu Rev Genet*, 30:465–506, 1996. 0066-4197 (Print) Journal Article Review.
- [86] A. Gragerov, E. Nudler, N. Komissarova, G. A. Gaitanaris, M. E. Gottesman, and V. Nikiforov. Cooperation of groel/groes and dnaK/dnaj heat shock proteins in preventing protein misfolding in escherichia coli. *Proc Natl Acad Sci U S A*, 89(21):10341–4, 1992.

- [87] L. M. Guzman, D. Belin, M. J. Carson, and J. Beckwith. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pbad promoter. *J Bacteriol*, 177(14):4121–30, 1995. Guzman, L M Belin, D Carson, M J Beckwith, J Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. United states Journal of bacteriology J Bacteriol. 1995 Jul;177(14):4121-30.
- [88] C Haas, A Y Hung, M Citron, D B Teplow, and D J Selkoe. beta-amyloid, protein processing and alzheimer's disease. *Arzneimittelforschung*, 45(3A):398–402, 1995.
- [89] F. U. Hartl. Molecular chaperones in cellular protein folding. *Nature*, 381(6583):571–9, 1996.
- [90] F. U. Hartl and M. Hayer-Hartl. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, 295(5561):1852–8, 2002.
- [91] C. S. Hayes, B. Bose, and R. T. Sauer. Stop codons preceded by rare arginine codons are efficient determinants of ssra tagging in escherichia coli. *Proc Natl Acad Sci U S A*, 99(6):3440–5, 2002.
- [92] A. Hayhurst and W. J. Harris. Escherichia coli skp chaperone coexpression improves solubility and phage display of single-chain antibody fragments. *Protein Expr Purif*, 15(3):336–43, 1999.
- [93] A. Helenius. How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell*, 5:253–265, 1994.
- [94] A. Helenius and M. Aebi. Intracellular functions of n-linked glycans. *Science*, 291(5512):2364–9, 2001.
- [95] T. Henrichs, N. Mikhaleva, C. Conz, E. Deuerling, D. Boyd, A. Zelazny, E. Bibi, N. Ban, and M. Ehrmann. Target-directed proteolysis at the ribosome. *Proc Natl Acad Sci U S A*, 102(12):4246–51, 2005.
- [96] A. Hoffmann, F. Merz, A. Rutkowska, B. Zachmann-Brand, E. Deuerling, and B. Bukau. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J Biol Chem*, 281(10):6539–45, 2006.
- [97] E. N. Houben, R. Zarivach, B. Oudega, and J. Luirink. Early encounters of a nascent membrane protein: specificity and timing of contacts inside and outside the ribosome. *J Cell Biol*, 170(1):27–35, 2005.

- [98] W. A. Houry. Chaperone-assisted protein folding in the cell cytoplasm. *Curr Protein Pept Sci*, 2(3):227–44, 2001. 1389-2037 (Print) Journal Article Review.
- [99] W. A. Houry, D. Frishman, C. Eckerskorn, F. Lottspeich, and F. U. Hartl. Identification of in vivo substrates of the chaperonin groel. *Nature*, 402(6758):147–54, 1999.
- [100] D. Huber, D. Boyd, Y. Xia, M. H. Olma, M. Gerstein, and J. Beckwith. Use of thioredoxin as a reporter to identify a subset of escherichia coli signal sequences that promote signal recognition particle-dependent translocation. *J Bacteriol*, 187(9):2983–91, 2005.
- [101] Z. Ignatova, B. Krishnan, J. P. Bombardier, A. M. Marcelino, J. Hong, and L. M. Gierasch. From the test tube to the cell: exploring the folding and aggregation of a beta-clam protein. *Biopolymers*, 88(2):157–63, 2007.
- [102] Zoya Ignatova and Lila M Gierasch. Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc Natl Acad Sci U S A*, 101(2):523–528, 2004.
- [103] B Imperiali and S E O'Connor. Effect of n-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol*, 3(6):643–649, 1999.
- [104] D. Japrun, S. Chusacultanachai, J. Yuvaniyama, P. Wilairat, and Y. Yuthavong. A simple dual selection for functionally active mutants of plasmodium falciparum dihydrofolate reductase with improved solubility. *Protein Eng Des Sel*, 18(10):457–64, 2005.
- [105] S. Jenni and N. Ban. The chemistry of protein synthesis and voyage through the ribosomal tunnel. *Curr Opin Struct Biol*, 13(2):212–9, 2003. 0959-440X (Print) Journal Article Research Support, Non-U.S. Gov't Review.
- [106] J. C. Joly, W. S. Leung, and J. R. Swartz. Overexpression of escherichia coli oxidoreductases increases recombinant insulin-like growth factor-i accumulation. *Proc Natl Acad Sci U S A*, 95(6):2773–7, 1998.
- [107] H. Kadokura, F. Katzen, and J. Beckwith. Protein disulfide bond formation in prokaryotes. *Annu Rev Biochem*, 72:111–35, 2003.

- [108] Tsutomu Kakuda and Victor J DiRita. Cj1496c encodes a campylobacter jejuni glycoprotein that influences invasion of human epithelial cells and colonization of the chick gastrointestinal tract. *Infect Immun*, 74(8):4715–4723, 2006.
- [109] J. F. Kane. Effects of rare codon clusters on high-level expression of heterologous proteins in escherichia coli. *Curr Opin Biotechnol*, 6(5):494–500, 1995.
- [110] J. F. Kane, B. N. Violand, D. F. Curran, N. R. Staten, K. L. Duffin, and G. Bogosian. Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of escherichia coli. *Nucleic Acids Res*, 20(24):6707–12, 1992.
- [111] R. B. Kapust and D. S. Waugh. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*, 8(8):1668–74, 1999. Kapust, R B Waugh, D S Research Support, U.S. Gov't, P.H.S. United states Protein science : a publication of the Protein Society Protein Sci. 1999 Aug;8(8):1668-74.
- [112] AV Karlyshev, P. Everest, D. Linton, S. Cawthraw, DG Newell, and BW Wren. The Campylobacter jejuni general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology*, 150(6):1957, 2004.
- [113] B. R. Kelemen, T. A. Klink, M. A. Behlke, S. R. Eubanks, P. A. Leland, and R. T. Raines. Hypersensitive substrate for ribonucleases. *Nucleic Acids Res*, 27(18):3696–701, 1999.
- [114] John Kelly, Harold Jarrell, Lorna Millar, Luc Tessier, Laura M Fiori, Peter C Lau, Brenda Allan, and Christine M Szymanski. Biosynthesis of the n-linked glycan in campylobacter jejuni and addition onto protein through block transfer. *J Bacteriol*, 188(7):2427–2434, 2006.
- [115] M. J. Kerner, D. J. Naylor, Y. Ishihama, T. Maier, H. C. Chang, A. P. Stines, C. Georgopoulos, D. Frishman, M. Hayer-Hartl, M. Mann, and F. U. Hartl. Proteome-wide analysis of chaperonin-dependent protein folding in escherichia coli. *Cell*, 122(2):209–20, 2005.
- [116] I Kheterpal, A Williams, C Murphy, B Bledsoe, and R Wetzel. Structural features of the abeta amyloid fibril elucidated by limited proteolysis. *Biochemistry*, 40(39):11757–11767, 2001.

- [117] C. H. Kim, Y. Oh, and T. H. Lee. Codon optimization for high-level expression of human erythropoietin (epo) in mammalian cells. *Gene*, 199(1-2):293–301, 1997.
- [118] J. Y. Kim, E. A. Fogarty, F. J. Lu, H. Zhu, G. D. Wheelock, L. A. Henderson, and M. P. DeLisa. Twin-arginine translocation of active human tissue plasminogen activator in escherichia coli. *Appl Environ Microbiol*, 71(12):8451–9, 2005.
- [119] Pan-Jun Kim, Dong-Yup Lee, and Hawoong Jeong. Centralized modularity of n-linked glycosylation pathways in mammalian cells. *PLoS One*, 4(10):e7317, 2009.
- [120] Sung Bae Kim, Takeaki Ozawa, Shigeaki Watanabe, and Yoshio Umezawa. High-throughput sensing and noninvasive imaging of protein nuclear transport by using reconstitution of split renilla luciferase. *Proc Natl Acad Sci U S A*, 101(32):11542–11547, 2004.
- [121] Woojin Kim, Yunkyong Kim, Jaeki Min, Dong Jin Kim, Young-Tae Chang, and Michael H Hecht. A high-throughput screen for compounds that inhibit aggregation of the alzheimer’s peptide. *ACS Chem Biol*, 1(7):461–469, 2006.
- [122] C. Kimchi-Sarfaty, J. M. Oh, I. W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar, and M. M. Gottesman. A “silent” polymorphism in the *mdr1* gene changes substrate specificity. *Science*, 315(5811):525–8, 2007.
- [123] S. Kirmizialtin, V. Ganesan, and D. E. Makarov. Translocation of a beta-hairpin-forming peptide through a cylindrical tunnel. *J Chem Phys*, 121(20):10268–77, 2004.
- [124] Holger Koch, Nico Grafe, Ralph Schiess, and Andreas Pluckthun. Direct selection of antibodies from complex libraries with the protein fragment complementation assay. *J Mol Biol*, 357(2):427–441, 2006.
- [125] A. Korostelev, S. Trakhanov, M. Laurberg, and H. F. Noller. Crystal structure of a 70s ribosome-trna complex reveals functional interactions and rearrangements. *Cell*, 126(6):1065–77, 2006.
- [126] Jan S Kostecki, Haiming Li, Raymond J Turner, and Matthew P DeLisa. Visualizing interactions along the escherichia coli twin-arginine translocation pathway using protein fragment complementation. *PLoS One*, 5(2):e9225, 2010.

- [127] L. Kotula and P. J. Curtis. Evaluation of foreign gene codon optimization in yeast: expression of a mouse ig kappa chain. *Biotechnology (N Y)*, 9(12):1386–9, 1991.
- [128] Michael Kowarik, Shin Numao, Mario F Feldman, Benjamin L Schulz, Nico Callewaert, Eva Kiermaier, Ina Catrein, and Markus Aebi. N-linked glycosylation of folded proteins by the bacterial oligosaccharyl-transferase. *Science*, 314(5802):1148–1150, 2006.
- [129] Michael Kowarik, N Martin Young, Shin Numao, Benjamin L Schulz, Isabelle Hug, Nico Callewaert, Dominic C Mills, David C Watson, Marcela Hernandez, John F Kelly, Michael Wacker, and Markus Aebi. Definition of the bacterial n-glycosylation site consensus sequence. *EMBO J*, 25(9):1957–1966, 2006.
- [130] M. Kraft, U. Knupfer, R. Wenderoth, P. Pietschmann, B. Hock, and U. Horn. An online monitoring system based on a synthetic sigma32-dependent tandem promoter for visualization of insoluble proteins in the cytoplasm of escherichia coli. *Appl Microbiol Biotechnol*, 2007.
- [131] Frederick J Krambeck and Michael J Betenbaugh. A mathematical model of n-linked glycosylation. *Biotechnol Bioeng*, 92(6):711–728, 2005.
- [132] G. Kramer, V. Ramachandiran, and B. Hardesty. Cotranslational folding—omnia mea mecum porto? *Int J Biochem Cell Biol*, 33(6):541–53, 2001.
- [133] G. Kramer, T. Rauch, W. Rist, S. Vorderwulbecke, H. Patzelt, A. Schulze-Specking, N. Ban, E. Deuerling, and B. Bukau. L23 protein functions as a chaperone docking site on the ribosome. *Nature*, 419(6903):171–4, 2002.
- [134] C A Kumamoto and J Beckwith. Evidence for specificity at an early step in protein export in escherichia coli. *J Bacteriol*, 163(1):267–274, 1985.
- [135] C A Kumamoto and P M Gannon. Effects of escherichia coli secB mutations on pre-maltose binding protein conformation and export kinetics. *J Biol Chem*, 263(23):11554–11558, 1988.
- [136] Y. Kurokawa, H. Yanagi, and T. Yura. Overproduction of bacterial protein disulfide isomerase (dsbc) and its modulator (dsbd) markedly enhances periplasmic production of human nerve growth factor in escherichia coli. *J Biol Chem*, 276(17):14393–9, 2001. 0021-9258 (Print) Journal Article.

- [137] R. Lafond, X. Zhan, and G. Georgiou. Screening and selection strategies for disulfide isomerase activity. *Methods Mol Biol*, 230:239–57, 2003. 1064-3745 (Print) Journal Article.
- [138] E. R. LaVallie, E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel, and J. M. McCoy. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the e. coli cytoplasm. *Biotechnology (N Y)*, 11(2):187–93, 1993. 0733-222X (Print) Journal Article.
- [139] P. A. Lee, D. Tullman-Ercek, and G. Georgiou. The bacterial twin-arginine translocation pathway. *Annu Rev Microbiol*, 60:373–95, 2006.
- [140] S. C. Lee and P. O. Olins. Effect of overproduction of heat shock chaperones groes1 and dnaK on human procollagenase production in escherichia coli. *J Biol Chem*, 267(5):2849–52, 1992.
- [141] S. A. Lesley, J. Graziano, C. Y. Cho, M. W. Knuth, and H. E. Klock. Gene expression response to misfolded protein as a screen for soluble recombinant protein. *Protein Eng*, 15(2):153–60, 2002.
- [142] Markus Liebscher, Gunther Jahreis, Christian Lucke, Susanne Grabley, Satish Raina, and Cordelia Schiene-Fischer. Fatty acyl benzamido antibacterials based on inhibition of dnaK-catalyzed protein folding. *J Biol Chem*, 282(7):4437–4446, 2007.
- [143] Hyung-Kwon Lim, Thomas J Mansell, Stephen W Linderman, Adam C Fisher, Michael R Dyson, and Matthew P DeLisa. Mining mammalian genomes for folding competent proteins using tat-dependent genetic selection in escherichia coli. *Protein Sci*, 18(12):2537–2549, 2009.
- [144] D. Linton, E. Allan, A.V. Karlyshev, A.D. Cronshaw, and B.W. Wren. Identification of N-acetylgalactosamine-containing glycoproteins PEB 3 and CgpA in *Campylobacter jejuni*. *Molecular Microbiology*, 43(2):497–508, 2002.
- [145] D. Linton, N. Dorrell, P.G. Hitchen, S. Amber, A.V. Karlyshev, H.R. Morris, A. Dell, M.A. Valvano, M. Aebi, and B.W. Wren. Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Molecular Microbiology*, 55(6):1695–1703, 2005.
- [146] J. W. Liu, Y. Boucher, H. W. Stokes, and D. L. Ollis. Improving protein solubility: the use of the escherichia coli dihydrofolate reductase gene as a fusion reporter. *Protein Expression Purif*, 47(1):258–63, 2006.

- [147] J. Lu and C. Deutsch. Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol*, 12(12):1123–9, 2005.
- [148] Chi-Hao Luan, Shihong Qiu, James B Finley, Mike Carson, Rita J Gray, Wenying Huang, David Johnson, Jun Tsao, Jerome Reboul, Philippe Vaglio, David E Hill, Marc Vidal, Lawrence J Delucas, and Ming Luo. High-throughput expression of *c. elegans* proteins. *Genome Res*, 14(10B):2102–2110, 2004.
- [149] K. Luby-Phelps. Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. *Int Rev Cytol*, 192:189–221, 2000. 0074-7696 (Print) Journal Article Research Support, U.S. Gov't, Non-P.H.S. Review.
- [150] J. Macfarlane and M. Muller. The functional integration of a polytopic membrane protein of *escherichia coli* is dependent on the bacterial signal-recognition particle. *Eur J Biochem*, 233(3):766–71, 1995.
- [151] T. J. Magliery and L. Regan. Combinatorial approaches to protein stability and structure. *Eur J Biochem*, 271(9):1595–608, 2004.
- [152] T. J. Magliery and L. Regan. Library approaches to biophysical problems. *Eur J Biochem*, 271(9):1593–4, 2004.
- [153] Thomas J Mansell, Stephen W Linderman, Adam C Fisher, and Matthew P DeLisa. A rapid protein folding assay for the bacterial periplasm. *Protein Sci*, 19(5):1079–1090, 2010.
- [154] W. Markland, B. L. Roberts, M. J. Saxena, S. K. Guterman, and R. C. Ladner. Design, construction and function of a multicopy display vector using fusions to the major coat protein of bacteriophage m13. *Gene*, 109(1):13–9, 1991.
- [155] P. Martineau, P. Jones, and G. Winter. Expression of an antibody fragment at high levels in the bacterial cytoplasm. *J Mol Biol*, 280(1):117–27, 1998. Martineau, P Jones, P Winter, G Research Support, Non-U.S. Gov't England Journal of molecular biology *J Mol Biol*. 1998 Jul 3;280(1):117-27.
- [156] K. L. Maxwell, A. K. Mittermaier, J. D. Forman-Kay, and A. R. Davidson. A simple in vivo assay for increased protein solubility. *Protein Sci*, 8(9):1908–11, 1999.

- [157] Y. Mazor, T. V. Blarcom, R. Mabry, B. L. Iverson, and G. Georgiou. Isolation of engineered, full-length antibodies from libraries expressed in escherichia coli. *Nat Biotechnol*, 25(5):563–5, 2007.
- [158] E. C. McCusker, S. E. Bane, M. A. O’Malley, and A. S. Robinson. Heterologous gpcr expression: a bottleneck to obtaining crystal structures. *Biotechnol Prog*, 23(3):540–7, 2007. McCusker, Emily C Bane, Steven E O’Malley, Michelle A Robinson, Anne Skaja P20 rr015588-076584/rr/ncrr P20-rr15588/rr/ncrr Research Support, N.I.H., Extramural Review United States Biotechnology progress *Biotechnol Prog*. 2007 May-Jun;23(3):540-7. Epub 2007 Mar 31.
- [159] K. E. McGinness, T. A. Baker, and R. T. Sauer. Engineering controllable protein degradation. *Mol Cell*, 22(5):701–7, 2006.
- [160] B. C. McNulty, G. B. Young, and G. J. Pielak. Macromolecular crowding in the escherichia coli periplasm maintains alpha-synuclein disorder. *J Mol Biol*, 355(5):893–7, 2006. 0022-2836 (Print) Journal Article Research Support, U.S. Gov’t, Non-P.H.S.
- [161] H. J. Meerman and G. Georgiou. Construction and characterization of a set of e. coli strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins. *Biotechnology (N Y)*, 12(11):1107–10, 1994. 0733-222X (Print) Journal Article Research Support, U.S. Gov’t, Non-P.H.S.
- [162] Tongpil Min, Masoud Vedadi, David C Watson, Gregory A Wasney, Christine Munger, Miroslaw Cygler, Allan Matte, and N Martin Young. Specificity of campylobacter jejuni adhesin peb3 for phosphates and structural differences among its ligand complexes. *Biochemistry*, 48(14):3057–3067, 2009.
- [163] B. Miroux and J. E. Walker. Over-production of proteins in escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol*, 260(3):289–98, 1996.
- [164] D. Missiakas, J. M. Betton, and S. Raina. New components of protein folding in extracytoplasmic compartments of escherichia coli sura, fkpa and skp/omph. *Mol Microbiol*, 21(4):871–84, 1996. 0950-382X (Print) Journal Article Research Support, Non-U.S. Gov’t.
- [165] D. Missiakas and S. Raina. Protein folding in the bacterial periplasm. *J Bacteriol*, 179(8):2465–71, 1997. 0021-9193 (Print) Journal Article Review.

- [166] M. Mujacic, K. W. Cooper, and F. Baneyx. Cold-inducible cloning vectors for low-temperature protein expression in escherichia coli: application to the production of a toxic and proteolytically sensitive fusion protein. *Gene*, 238(2):325–32, 1999.
- [167] Axel Muller, Gavin H Thomas, Richard Horler, James A Brannigan, Elena Blagova, Vladimir M Levdikov, Mark J Fogg, Keith S Wilson, and Anthony J Wilkinson. An atp-binding cassette-type cysteine transporter in campylobacter jejuni inferred from the structure of an extracytoplasmic solute receptor protein. *Mol Microbiol*, 57(1):143–155, 2005.
- [168] C. Murphy, W. Prinz, M. Pohlschroder, A. Derman, and J. Beckwith. Essential features of the pathway for protein translocation across the escherichia coli cytoplasmic membrane. *Cold Spring Harbor Symp Quant Biol*, 60:277–83, 1995. 0091-7451 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [169] Y. Nakamura, T. Gojobori, and T. Ikemura. Codon usage tabulated from international dna sequence databases: status for the year 2000. *Nucleic Acids Res*, 28(1):292, 2000.
- [170] H. Nakatogawa, A. Murakami, and K. Ito. Control of secA and secM translation by protein secretion. *Curr Opin Microbiol*, 7(2):145–50, 2004.
- [171] H. C. Neu and L. A. Heppel. The release of enzymes from escherichia coli by osmotic shock and during the formation of spheroplasts. *J Biol Chem*, 240(9):3685–92, 1965. 0021-9258 (Print) In Vitro Journal Article.
- [172] K. Nishihara, M. Kanemori, M. Kitagawa, H. Yanagi, and T. Yura. Chaperone coexpression plasmids: differential and synergistic roles of dnaK-dnaJ-grpE and groEL-groES in assisting folding of an allergen of japanese cedar pollen, cryj2, in escherichia coli. *Appl Environ Microbiol*, 64(5):1694–9, 1998.
- [173] K. Nishihara, M. Kanemori, H. Yanagi, and T. Yura. Overexpression of trigger factor prevents aggregation of recombinant proteins in escherichia coli. *Appl Environ Microbiol*, 66(3):884–889, 2000.
- [174] Mihai Nita-Lazar, Michael Wacker, Belinda Schegg, Saba Amber, and Markus Aebi. The n-x-s/t consensus sequence is required but not sufficient for bacterial n-linked protein glycosylation. *Glycobiology*, 15(4):361–367, 2005.

- [175] N.B. Olivier, M.M. Chen, J.R. Behr, and B. Imperiali. In vitro biosynthesis of UDP-N, N'-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system. *Biochemistry*, 45(45):13659–13669, 2006.
- [176] Jagroop Pandhal, Saw Yen Ow, Josselin Noirel, and Phillip C Wright. Improving n-glycosylation efficiency in *Escherichia coli* using shotgun proteomics, metabolic network analysis, and selective reaction monitoring. *Biotechnol Bioeng*, 108(4):902–912, 2011.
- [177] M. Paschke and W. Hohne. A twin-arginine translocation (tat)-mediated phage display system. *Gene*, 350(1):79–88, 2005.
- [178] J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger, and G. S. Waldo. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol*, 24(1):79–88, 2006.
- [179] Jean-Denis Pedelacq, Emily Piltch, Elaine C Liong, Joel Berendzen, Chang-Yub Kim, Beom-Seop Rho, Min S Park, Thomas C Terwilliger, and Geoffrey S Waldo. Engineering soluble proteins for structural genomics. *Nat Biotechnol*, 20(9):927–932, 2002.
- [180] J. N. Pelletier, F. X. Campbell-Valois, and S. W. Michnick. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc Natl Acad Sci U S A*, 95(21):12141–6, 1998.
- [181] J. Perez-Perez, J. L. Barbero, G. Marquez, and J. Gutierrez. Different prla proteins increase the efficiency of periplasmic production of human interleukin-6 in *Escherichia coli*. *J Biotechnol*, 49(1-3):245–7, 1996.
- [182] R. Perez-Rodriguez, A. C. Fisher, J. D. Perlmutter, M. G. Hicks, A. Chanal, C. L. Santini, L. F. Wu, T. Palmer, and M. P. DeLisa. An essential role for the DnaK molecular chaperone in stabilizing over-expressed substrate proteins of the bacterial twin-arginine translocation pathway. *J Mol Biol*, 367(3):715–30, 2007. 0022-2836 (Print) Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.
- [183] C Pernelle, F F Clerc, C Dureuil, L Bracco, and B Tocque. An efficient screening assay for the rapid and precise determination of affinities between leucine zipper domains. *Biochemistry*, 32(43):11682–11687, 1993.

- [184] P. Philibert and P. Martineau. Directed evolution of single-chain fv for cytoplasmic expression using the beta-galactosidase complementation assay results in proteins highly susceptible to protease degradation and aggregation. *Microb Cell Fact*, 3(1):16, 2004.
- [185] T. Powers and P. Walter. Co-translational protein targeting catalyzed by the escherichia coli signal recognition particle and its receptor. *EMBO J*, 16(16):4880–6, 1997. Powers, T Walter, P Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. England The EMBO journal EMBO J. 1997 Aug 15;16(16):4880-6.
- [186] W. A. Prinz, F. Aslund, A. Holmgren, and J. Beckwith. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the escherichia coli cytoplasm. *J Biol Chem*, 272(25):15661–15667, 1997.
- [187] W. A. Prinz, C. Spiess, M. Ehrmann, C. Schierle, and J. Beckwith. Targeting of signal sequenceless proteins for export in escherichia coli with altered protein translocase. *EMBO J*, 15(19):5209–17, 1996.
- [188] A. P. Pugsley. The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev*, 57(1):50–108, 1993.
- [189] LL Randall and SJ Hardy. Correlation of competence for export with lack of tertiary structure of the mature species: a study in vivo of maltose-binding protein in E. coli. *Cell*, 46(6):921–8, 1986.
- [190] Jennifer L Reed, Iman Famili, Ines Thiele, and Bernhard O Palsson. Towards multidimensional genome annotation. *Nat Rev Genet*, 7(2):130–141, 2006.
- [191] Ingrid Remy, F X Campbell-Valois, and Stephen W Michnick. Detection of protein-protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. *Nat Protoc*, 2(9):2120–2125, 2007.
- [192] B. Ribnicky, T. Van Blarcom, and G. Georgiou. A scfv antibody mutant isolated in a genetic screen for improved export via the twin arginine transporter pathway exhibits faster folding. *J Mol Biol*, 369(3):631–9, 2007.
- [193] S. K. Rockenbach, M. J. Dupuis, T. W. Pitts, C. K. Marschke, and C. S. Tomich. Secretion of active truncated cd4 into escherichia coli periplasm. *Appl Microbiol Biotechnol*, 35(1):32–7, 1991.

- [194] H. S. Rollema, O. P. Kuipers, P. Both, W. M. de Vos, and R. J. Siezen. Improvement of solubility and stability of the antimicrobial peptide nisin by protein engineering. *Appl Environ Microbiol*, 61(8):2873–2878, 1995.
- [195] A. H. Rosenberg, E. Goldman, J. J. Dunn, F. W. Studier, and G. Zubay. Effects of consecutive agg codons on translation in escherichia coli, demonstrated with a versatile codon test system. *J Bacteriol*, 175(3):716–22, 1993.
- [196] D. Sachdev and J. M. Chirgwin. Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Expression Purif*, 12(1):122–132, 1998.
- [197] Joseph Sambrook and David W. Russell. *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 3rd edition, 2001. 00064380 Joseph Sambrook, David W. Russell. ill. ; 28 cm. Includes bibliographical references and index.
- [198] G. Schatz and B. Dobberstein. Common principles of protein translocation across membranes. *Science*, 271(5255):1519–26, 1996.
- [199] C. F. Schierle, M. Berkmen, D. Huber, C. Kumamoto, D. Boyd, and J. Beckwith. The dsba signal sequence directs efficient, cotranslational export of passenger proteins to the escherichia coli periplasm via the signal recognition particle pathway. *J Bacteriol*, 185(19):5706–13, 2003. Schierle, Clark F Berkmen, Mehmet Huber, Damon Kumamoto, Carol Boyd, Dana Beckwith, Jon Ai38591/ai/niaid Gm41883/gm/nigms Research Support, U.S. Gov't, P.H.S. United States Journal of bacteriology *J Bacteriol*. 2003 Oct;185(19):5706-13.
- [200] I.C. Schoenhofen, D.J. McNally, E. Vinogradov, D. Whitfield, N.M. Young, S. Dick, W.W. Wakarchuk, J.R. Brisson, and S.M. Logan. Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*. *Journal of Biological Chemistry*, 281(2):723, 2006.
- [201] Flavio Schwarz, Wei Huang, Cishan Li, Benjamin L Schulz, Christian Lizak, Alessandro Palumbo, Shin Numao, Dario Neri, Markus Aebi, and Lai-Xi Wang. A combined method for producing homogeneous glycoproteins with eukaryotic n-glycosylation. *Nat Chem Biol*, 6(4):264–266, 2010.
- [202] Flavio Schwarz, Christian Lizak, Yao-Yun Fan, Susanna Fleurkens, Michael Kowarik, and Markus Aebi. Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology*, 21(1):45–54, 2011.

- [203] Nichollas E Scott, Benjamin L Parker, Angela M Connolly, Jana Paulech, Alistair V G Edwards, Ben Crossett, Linda Falconer, Daniel Kolarich, Steven P Djordjevic, Peter Hojrup, Nicolle H Packer, Martin R Larsen, and Stuart J Cordwell. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by cid, higher energy collisional dissociation, and electron transfer dissociation ms applied to the n-linked glycoproteome of campylobacter jejuni. *Mol Cell Proteomics*, 10(2):M000031–MCP201, 2011.
- [204] L. Segatori, P. J. Paukstelis, H. F. Gilbert, and G. Georgiou. Engineered dsbc chimeras catalyze both protein oxidation and disulfide-bond isomerization in escherichia coli: Reconciling two competing pathways. *Proc Natl Acad Sci U S A*, 101(27):10018–23, 2004.
- [205] M. Selmer, C. M. Dunham, F. V. th Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, and V. Ramakrishnan. Structure of the 70s ribosome complexed with mrna and trna. *Science*, 313(5795):1935–42, 2006.
- [206] Guennadi Sezonov, Daniele Joseleau-Petit, and Richard D’Ari. Escherichia coli physiology in luria-bertani broth. *J Bacteriol*, 189(23):8746–8749, 2007.
- [207] C A Sheppard, E E Trimmer, and R G Matthews. Purification and properties of nadh-dependent 5, 10-methylenetetrahydrofolate reductase (metf) from escherichia coli. *J Bacteriol*, 181(3):718–725, 1999.
- [208] L. C. Simmons, D. Reilly, L. Klimowski, T. S. Raju, G. Meng, P. Sims, K. Hong, R. L. Shields, L. A. Damico, P. Rancatore, and D. G. Yansura. Expression of full-length immunoglobulins in escherichia coli: rapid and efficient production of aglycosylated antibodies. *J Immunol Methods*, 263(1-2):133–47, 2002.
- [209] D. B. Smith and K. S. Johnson. Single-step purification of polypeptides expressed in escherichia coli as fusions with glutathione s-transferase. *Gene*, 67(1):31–40, 1988. 0378-1119 (Print) Journal Article Research Support, Non-U.S. Gov’t.
- [210] G. P. Smith. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705):1315–7, 1985.
- [211] M. Sone, S. Kishigami, T. Yoshihisa, and K. Ito. Roles of disulfide bonds in bacterial alkaline phosphatase. *J Biol Chem*, 272(10):6174–8, 1997. Sone,

- M Kishigami, S Yoshihisa, T Ito, K Research Support, Non-U.S. Gov't United states The Journal of biological chemistry J Biol Chem. 1997 Mar 7;272(10):6174-8.
- [212] H. P. Sorensen, J. E. Kristensen, H. U. Sperling-Petersen, and K. K. Mortensen. Soluble expression of aggregating proteins by covalent coupling to the ribosome. *Biochem Biophys Res Commun*, 319(3):715–9, 2004.
- [213] M. A. Sorensen, C. G. Kurland, and S. Pedersen. Codon usage determines translation rate in escherichia coli. *J Mol Biol*, 207(2):365–77, 1989.
- [214] R. A. Spanjaard and J. van Duin. Translation of the sequence agg-agg yields 50 *Proc Natl Acad Sci U S A*, 85(21):7967–71, 1988.
- [215] D. Steiner, P. Forrer, M. T. Stumpp, and A. Pluckthun. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat Biotechnol*, 24(7):823–31, 2006.
- [216] E. M. Strauch and G. Georgiou. A bacterial two-hybrid system based on the twin-arginine transporter pathway of e. coli. *Protein Sci*, 16(5):1001–8, 2007.
- [217] R. A. Stuart and W. Neupert. Making membranes in bacteria. *Nature*, 406(6796):575, 577, 2000.
- [218] M. Suzuki, J. Zhang, M. Liu, N. A. Woychik, and M. Inouye. Single protein production in living cells facilitated by an mrna interferase. *Mol Cell*, 18(2):253–61, 2005. 1097-2765 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [219] J. R. Swartz. Advances in escherichia coli production of therapeutic proteins. *Curr. Opin. Biotechnol*, 12(2), 2001.
- [220] C. M. Szymanski and B. W. Wren. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol*, 3(3):225–37, 2005.
- [221] C.M. Szymanski, S.M. Logan, D. Linton, and B.W. Wren. Campylobacter-a tale of two protein glycosylation systems. *Trends in microbiology*, 11(5):233–238, 2003.
- [222] C.M. Szymanski, R. Yao, C.P. Ewing, T.J. Trust, and P. Guerry. Evidence for

- a system of general protein glycosylation in *Campylobacter jejuni*. *Molecular Microbiology*, 32(5):1022–1030, 1999.
- [223] Xuehai Tan, Pam Sanders, Jack Jr Bolado, and Mike Whitney. Integration of g-protein coupled receptor signaling pathways for activation of a transcription factor (*egr-3*). *Genomics Proteomics Bioinformatics*, 1(3):173–179, 2003.
- [224] P. Thammawong, W. Kasinrerak, R. J. Turner, and C. Tayapiwatana. Twin-arginine signal peptide attributes effective display of cd147 to filamentous phage. *Appl Microbiol Biotechnol*, 69(6):697–703, 2006.
- [225] T. A. Thanaraj and P. Argos. Ribosome-mediated translational pause and protein domain organization. *Protein Sci*, 5(8):1594–612, 1996.
- [226] J. G. Thomas and F. Baneyx. Protein misfolding and inclusion body formation in recombinant escherichia coli cells overexpressing heat-shock proteins. *J Biol Chem*, 271(19):11141–7, 1996.
- [227] J. G. Thomas and F. Baneyx. Divergent effects of chaperone overexpression and ethanol supplementation on inclusion body formation in recombinant escherichia coli. *Protein Expression Purif*, 11(3):289–96, 1997. 1046-5928 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.
- [228] T Tomoyasu, A Mogk, H Langen, P Goloubinoff, and B Bukau. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the escherichia coli cytosol. *Mol Microbiol*, 40(2):397–413, 2001.
- [229] J.M. Troutman and B. Imperiali. *Campylobacter jejuni* PglH Is a Single Active Site Processive Polymerase that Utilizes Product Inhibition to Limit Sequential Glycosyl Transfer Reactions. *Biochemistry*, 48(12):2807–2816, 2009.
- [230] K. Tsumoto, M. Umetsu, I. Kumagai, D. Ejima, and T. Arakawa. Solubilization of active green fluorescent protein from insoluble particles by guanidine and arginine. *Biochem Biophys Res Commun*, 312(4):1383–6, 2003.
- [231] D. Tullman-Ercek, M. P. DeLisa, Y. Kawarasaki, P. Iranpour, B. Ribnicky, T. Palmer, and G. Georgiou. Export pathway selectivity of escherichia coli twin arginine translocation signal peptides. *J Biol Chem*, 282(11):8309–16, 2007.

- [232] R. S. Ullers, E. N. Houben, A. Raine, C. M. ten Hagen-Jongman, M. Ehrenberg, J. Brunner, B. Oudega, N. Harms, and J. Luirink. 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(4):679–84, 2003.
- [233] P. UMAFIA and JE BAILEY. A mathematical model of N-linked glycoform biosynthesis. *Biotechnology and bioengineering*, 55(6):890–908, 1997.
- [234] Q.A. Valent, P.A. Scotti, S. High, J.W.L. de Gier, G. von Heijne, G. Lentzen, W. Wintermeyer, B. Oudega, and J. Luirink. The Escherichia coli SRP and SecB targeting pathways converge at the translocon. *EMBO Journal*, 17(9):2504–2512, 1998.
- [235] B. Van den Berg, Jr. Clemons, W. M., I. Collinson, Y. Modis, E. Hartmann, S. C. Harrison, and T. A. Rapoport. X-ray structure of a protein-conducting channel. *Nature*, 427(6969):36–44, 2004. 1476-4687 (Electronic) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S.
- [236] N. Varadarajan, J. Gam, M. J. Olsen, G. Georgiou, and B. L. Iverson. Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc Natl Acad Sci U S A*, 102(19):6855–60, 2005.
- [237] P. M. Viallet and T. Vo-Dinh. Monitoring intracellular proteins using fluorescence techniques: from protein synthesis and localization to activity. *Curr Protein Pept Sci*, 4(5):375–88, 2003. Viallet, Pierre M Vo-Dinh, Tuan Research Support, Non-U.S. Gov't Review Netherlands Current protein peptide science *Curr Protein Pept Sci*. 2003 Oct;4(5):375-88.
- [238] Antonio Villaverde and M Mar Carrio. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol Lett*, 25(17):1385–1395, 2003.
- [239] M. Wacker, D. Linton, P.G. Hitchen, M. Nita-Lazar, S.M. Haslam, S.J. North, M. Panico, H.R. Morris, A. Dell, B.W. Wren, et al. N-Linked Glycosylation in *Campylobacter jejuni* and Its Functional Transfer into *E. coli*. *Science*, 298(5599):1790–1793, 2002.
- [240] G. S. Waldo, B. M. Standish, J. Berendzen, and T. C. Terwilliger. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol*, 17(7):691–5, 1999.

- [241] G.S. Waldo. Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol*, 7:33–38, 2003.
- [242] Harris H Wang, Farren J Isaacs, Peter A Carr, Zachary Z Sun, George Xu, Craig R Forest, and George M Church. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257):894–898, 2009.
- [243] J. D. Wang, C. Herman, K. A. Tipton, C. A. Gross, and J. S. Weissman. Directed evolution of substrate-optimized groel/s chaperonins. *Cell*, 111(7):1027–39, 2002. 0092-8674 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S.
- [244] D. Waraho and M.P. DeLisa. A versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proc Natl Acad Sci U S A*, page submitted, 2007.
- [245] Dujduan Waraho and Matthew P DeLisa. Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proc Natl Acad Sci U S A*, 106(10):3692–3697, 2009.
- [246] Joseph R Warner, Philippa J Reeder, Anis Karimpour-Fard, Lauren B A Woodruff, and Ryan T Gill. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol*, 28(8):856–862, 2010.
- [247] E. Weerapana, K.J. Glover, M.M. Chen, and B. Imperiali. Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *Journal of the American Chemical Society*, 127(40):13766–13767, 2005.
- [248] Eranthie Weerapana and Barbara Imperiali. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology*, 16(6):91R–101R, 2006.
- [249] Tom Wehrman, Benjamin Kleaveland, Jeng-Horng Her, Robert F Balint, and Helen M Blau. Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. *Proc Natl Acad Sci U S A*, 99(6):3469–3474, 2002.

- [250] W. C. Wigley, R. D. Stidham, N. M. Smith, J. F. Hunt, and P. J. Thomas. Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nat Biotechnol*, 19(2):131–6, 2001.
- [251] S. Wildt and T.U. Gerngross. The humanization of N-glycosylation pathways in yeast. *Nature Reviews Microbiology*, 3(2):119–128, 2005.
- [252] C. A. Woolhead, P. J. McCormick, and A. E. Johnson. Nascent membrane and secretory proteins differ in fret-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell*, 116(5):725–36, 2004.
- [253] C. Wurth, N.K. Guimard, and M.H. Hecht. Mutations that reduce aggregation of the Alzheimer’s AB42 peptide: An unbiased search for the sequence determinants of AB amyloidogenesis. *J. Mol. Biol*, 319:1279–1290, 2002.
- [254] Agnieszka Wyszynska, Joanna Zycka, Renata Godlewska, and Elzbieta K Jagusztyn-Krynicka. The campylobacter jejuni/coli cjaa (cj0982c) gene encodes an n-glycosylated lipoprotein localized in the inner membrane. *Curr Microbiol*, 57(3):181–188, 2008.
- [255] J. C. Young, V. R. Agashe, K. Siegers, and F. U. Hartl. Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol*, 5(10):781–91, 2004.
- [256] M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller. Crystal structure of the ribosome at 5.5 a resolution. *Science*, 292(5518):883–96, 2001.
- [257] Y. Zhang, J. Zhang, H. Hara, I. Kato, and M. Inouye. Insights into the mrna cleavage mechanism by mazf, an mrna interferase. *J Biol Chem*, 280(5):3143–50, 2005.
- [258] Y. Zhang, J. Zhang, K. P. Hoeflich, M. Ikura, G. Qing, and M. Inouye. Mazf cleaves cellular mRNAs specifically at aca to block protein synthesis in escherichia coli. *Mol Cell*, 12(4):913–23, 2003.
- [259] G. Ziv, G. Haran, and D. Thirumalai. Ribosome exit tunnel can entropically stabilize alpha-helices. *Proc Natl Acad Sci U S A*, 102(52):18956–61, 2005. 0027-8424 (Print) Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, Non-P.H.S.