EXPANDING THE CAPABILITIES OF A BACTERIAL QUALITY CONTROL MECHANISM FOR ENGINEERING ENZYMES

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

Jason Thomas Boock

January 2015

EXPANDING THE CAPABILITIES OF A BACTERIAL QUALITY CONTROL MECHANISM FOR ENGINEERING ENZYMES

Jason Thomas Boock, Ph.D.

Cornell University 2015

In this study, we have repurposed the intrinsic quality control of the twin-arginine translocation (Tat) pathway to enhance the traits of target proteins using directed evolution. As a proof-of-concept, we applied this method to the endocellulase Cel5A from the fungal plant pathogen *Fusarium graminearum* to improve its production and function. Our approach is based on a novel two-tiered genetic selection and screening method. First, a Tat-based genetic selection is applied that links protein translocation with resistance to beta-lactam antibiotics. Since the quality control mechanism of the Tat pathway only permits the export of folded proteins, this genetic selection allows for the rapid and high-throughput isolation of well-folded, stable Cel5A library members while eliminating those that are poorly folded. A second screening step is imposed to ensure that the proteins that pass the Tat quality control retain high activity. For Cel5A, this involves a screen of enzyme activity using the soluble cellulose substrate carboxymethyl cellulose. Following two iterations through our dual selection and screen, we isolated a Cel5A variant whose production is increased 30-fold over the parent enzyme. The gain in production is achieved without any loss in activity on soluble or insoluble cellulose substrates, underscoring the value in this two-step evolution approach. Further, we have characterized several of the improved variants to

begin determining which biophysical properties are selected through the directed evolution process. Additionally, we have discovered the chaperone-like activity of a component of the Tat pathway. The results of these experiments are helping to shed light on the poorly understood quality control mechanism and will guide future protein engineering attempts that exploit this pathway.

BIOGRAPHICAL SKETCH

I was born in the small town of Sugarloaf, Pennsylvania where I lived on a farm with my parents and younger sister. I began my journey as a scientist in high school, presenting on research topics ranging from determining friction coefficients for tires to killing termites with antibiotics. Outside of school, I was an active member of our local Boy Scout troop where I learned more than just how to tie knots and build fires. My desire to perform research led me to attend the Johns Hopkins University for my undergraduate studies. In my sophomore year I joined a protein engineering lab where I focused on modifying an artificial protein switch and studying its mechanism of allostery. My interest in science and teaching brought me to graduate school at Cornell University, and I continue to learn more about each every day. There I also discovered my passion for mentorship and high school outreach. Over the past two seasons, I compiled an 11-3 record as the starting pitcher for the department summer softball team.

This work is dedicated to everyone listed in the acknowledgements.

Someone once told me that acknowledgements are free, so list everyone.

This couldn't be further from the truth.

It is only through your assistance, advice, love, dedication, and willingness to give up your precious time that this work lives and I made it through this journey.

And Cornell pays for the extra ink.

ACKNOLEDGEMENTS

To Matt, for graciously providing guidance, support, passion, knowledge and mentorship. For bringing the team together, creating a successful laboratory environment, and diligently pushing me along.

To Rob, for showing me the ropes, having a great work ethic, and answering all my inane questions.

To Brian, for teaching me everything about cellulases and being a great partner.

To Alyse, for always being there regardless of the request, the great discussions, and your approach to life.

To Linxiao, for being a great friend, peer and all around fantastic person, not to mention proofreading all my emails.

To Tom, for being my “scientific older brother” and the late night conversations about “research”.

To Taylor, for always being willing to lend a hand and your excellent ideas.

To May, for showing me how much I enjoy mentorship and making me jealous of any student I will have in the future.

To Aravind, for your passionate outlook on research, never being afraid to ask any question, and the surprise parties.

To Rick, for all your excellent questions, unique tales, and never-say-never attitude.

To Adam, for being an inspiration and your approach to science.

To Christine, for your friendship and always being there for each other.

To Kay, for patiently learning mentorship with me, all your hard work and your

independence.

To Jessica, for sticking with the project even when results were hard to come by.

To Maria, Rachel I, Rachel II, and Aidan, for being great students and always showing up to lab with a smile on your face.

To Amy and Jenny, for all the conversations about academics and graduate school.

To Sean, for your unique ideas and passionately following them.

To Dario, for teaching me to think outside the box.

To Charles, for your advice and methodical approach to science.

To M-P, for your diligent and unwavering attitude toward research.

To Erin, for the awesome cake and happy personality.

To Didi, Cassie, Jan, Matt (Grey), Jaun, Emily, Sai, Anne, Dr. Lee, Pao, Ritz, Mark, Eda, Xialou, Morgan, Cameron, Yi, Tommy and Lydia for being great partners in lab.

To Shelby, for being our department Mom and giving unending support.

To Carol, Sue and secretaries, for answering all my emails and providing the foundation for all that we do.

To Jeff, for being a great person to talk to and always willing to share past experiences.

To Abe, for guiding me to find my passion in teaching and showing me what it is like to be a scientific leader.

To Paulette, for sculpting the GK12 experience and being a role model.

To Dave and Hening, for being on my committee and the great feedback

To Cynthia, for all your FPLC wisdom.

To David, for proving that scholarship is still important in higher education.

To Sameer, for changing how I view the world and education.

To Larry, for your take on science, mentorship and the community.

To Al, for casual conversations and giving me the opportunity to lead.

To Jim, for talking about everything.

To Wayne and David, for shaping the person I am and all your wisdom.

To Marc, for providing me with a research problem and showing me what academic research is all about.

To the ChemBEasts, ChEGSA, and Cornell Hockey Crew, for getting me outside lab.

To Ithaca, for all your beautiful people, I know more about myself because of them.

To the NSF, USDA, and NYSTAR, for graciously funding this process.

To my parents, sister and uncles, for getting me started on this journey, teaching me a strong work ethic, and never accepting anything but my best.

To Nate, Jeremy, Joey, Martin, and Holly, for being the best friends a guy could ask for and always willing to share a story.

To Kristen, for listening to all my complaints, staying positive, being patient and your everlasting love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

a.u.	Arbitrary units
AmiA	N-acetylmuramyl-L-alanine amidase A
AmiC	N-acetylmuramyl-L-alanine amidase C
aTc	Anhydrous tetracycline
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BioSAXS	Biological small angle X-ray scattering
Bla	TEM1 β-lactamase
BMCC	Bacterial microcrystalline cellulose
BSA	Bovine serum albumin
Carb	Carbenicillin antibiotic
Cat	Chloramphenicol acetyl transferase
CAzy	Carbohydrate-active enzyme
CBM	Carbohydrate binding module
CBU	Cellobiase units
CD	Catalytic domain
Cel5A	Cellulase family 5 protein A
CFU	Colony forming unit
Cm	Chloramphenicol antibiotic
CMC	Carboxymethyl cellulose
cv	Column volumes
CWDE	Cell wall degrading enzymes
DmsA	Dimethyl sulfoxide reductase protein A
DmsD	Dimethyl sulfoxide reductase chaperone D
DNA	Deoxyribonucleic acid
DNS	3,5-dinitrosalicylic acid
DsbC	Disulfide bond isomerase C
E. coli	*Escherichia coli*
EC	Enzyme classification
EDTA	Ethylenediaminetetraacetic acid
Em	Emission wavelength
Eno	Enolase
Ex	Excitation wavelength
F. graminearum	*Fusarium graminearum*
FL	Full-length
FPLC	Fast protein liquid chromatography
FRET	Förester resonance energy transfer
Fv	Variable domain
GFP	Green fluorescent protein
GH	Glycoside hydrolase
GroEL	60 kDa chaperonin protein
HCl	Hydrochloric acid

IMAC	Immobilized metal affinity chromatography
IPTG	Isoprpyl β-D-1-thiogalactopyranoside
k_{cat}	Michaelis-Menten kinetic rate constant
k_D	Dissociation constan
kDa	10^3 daltons (or 10^3 grams/mol)
K_m	Michaelis constant
LB	Luria-Bertani
mAU	milli-absorbance units
MBP	Maltose binding protein
Met_f	Methionine start codon
MW	Molecular weight
PAGE	Protein acrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pH	Hydrogen ion concentration
PhoA	Alkaline phosphatase
PMF	Proton motive force
ProtK	Proteinase K
RNA	Ribonucleic acid
RNC	Ribosome nascent chai
Sec	General secretion pathway
SecA	Secretion pathway protein A
scFv	Single-chain variable fragment
SDS	Sodium dodecyl sulfate
SPase	Signal peptidase
SRP	Signal recognition particle
ss	Signal sequence
SrtA	*Staphylococcus aureus* sortaseA
T. reesei	*Trichoderma reesei* (formally known as *Hypocrea jecorina*)
Tat	Twin-arginine translocation
TatA	Twin-arginine translocation pathway protein A
TatB	Twin-arginine translocation pathway protein B
TatC	Twin-arginine translocation pathway protein C
TatD	Twin-arginine translocation pathway protein D
TB	Terrific broth
TorA	Trimethhylamine-*N*-oxide reductase protein A
TorD	Trimethhylamine-*N*-oxide reductase chaperone D
V_e	Elution volume
V_o	Void volume
v_o	Initial rate
wt	Wild-type

CHAPTER 1

INTRODUCTION: THE TWIN-ARGININE TRANSLOCATION PATHWAY AND QUALITY CONTROL BASED PROTEIN ENGINEERING

Heterologous protein production in E. coli

Since Crick and Watson unraveled the universal genetic code in 1953, scientists have dreamed of endowing the characteristics of one organism to another through genetic transfer. Heterologous protein expression has since become commonplace in laboratories; permitting the characterization of proteins across species, incorporation of pathways to create novel compounds, production of potent pharmaceuticals, and construction of proteins never before seen in nature [1]. Entire biotechnology fields are built around recombinant DNA technology including, protein crystallography, synthetic biology, and metabolic engineering, to name a few. Arguably, the first recombinant pharmaceutical, Humulin, is still the biggest success of the heterologous protein era, enabling the treatment of diabetes without needing to collect animal pancreases [2]. Today, many of the top grossing pharmaceuticals are recombinant antibodies, garnering 24.6 billion dollars in sales in 2012 alone [3]. Heterologous protein production has influenced far more than the biotech market and without it biological systems would be far less mature.

The gram-negative enteric bacterium *Escherichia coli* remains the preferred host for heterologous protein production [4]. As of this writing, “Recombinant protein expression in *E. coli*” yields over 106,000 articles when searched using Web of Science.

E. coli has become the workhorse of biotechnology due to its well understood and easily manipulated genetics, low culture cost, fast doubling times and ability to grow to high density in fermentations [5-7]. Strain engineering has been progressively applied to this bacterium to enhance protein titers through incorporation of chaperones, deletion of proteases, and modification of natural pathways [1, 8]. Optimization of media formulation, plasmid combinations, RNA stability, promoter design, and culture conditions have additionally increased protein yields, and the ability to perform chemistries inside cells of which synthetic chemists can only dream [6].

For all the success stories of recombinant protein production in *E. coli*, there are still numerous drawbacks to using this organism to produce proteins of interest. Misfolding of proteins is a major concern due to high intracellular protein concentration and an incorrect balance of slow non-native protein-folding kinetics with the rapid ribosome translation speed in *E. coli*, often resulting in aggregation of proteins in inclusion bodies [8]. Co-translational folding of protein fusions is problematic due to nascent polypeptide chains prematurely folding on top of each other [9, 10]. Chaperones or cofactors that direct protein-folding pathways may be absent for heterologous proteins, leading to inappropriately folded, inactive proteins [11]. Post-translational modifications, such as glycosylation or phosphorylation, are naturally absent and require incorporation of other recombinant enzyme pathways [1]. Disulfide bonding occurs in the periplasm of *E. coli*, but requires the energy intensive and low capacity general secretory pathway to translocate proteins there [12, 13]. Lastly, few extracellular protein secretion pathways have been discovered in *E. coli* [14],

necessitating intracellular production followed by cell lysis to recover protein products. These disadvantages and better genetic tools have enabled diverse heterologous protein expression hosts to become widely used in biotechnology [15-17]; however, for the past ten years, over 75% of heterologous proteins continue to be produced in *E. coli* [4].

The twin-arginine translocation pathway

Most gram-negative bacteria have two protein translocation pathways to export proteins produced in the cytoplasm into the periplasmic space (**Figure 1.1a**) [18]. The general secretion pathway (Sec) is ubiquitous among all forms of life, is essential, and is the preferred pathway for translocation of most proteins. Like translocation into the endoplasmic reticulum of eukaryotic cells, the Sec machinery in gram-negative bacteria transports unfolded polypeptides into the periplasm, where the folds amid periplasmic chaperones, including disulfide bond creating proteins [19, 20]. The twin-arginine translocation (Tat) system is a second pathway for exporting proteins from the cytoplasm to periplasm. Unlike the Sec pathway, the Tat system transports only fully folded proteins and does so in an ATP-independent manner [21]. This pathway was originally discovered in plant thylakoid membranes over 15 years ago and later found to be conserved in many prokaryotes [22, 23].

The Tat system is eponymously named from the presence of a twin arginine dipeptide in the leader signal sequence of target proteins [21, 24, 25]. In bacteria,

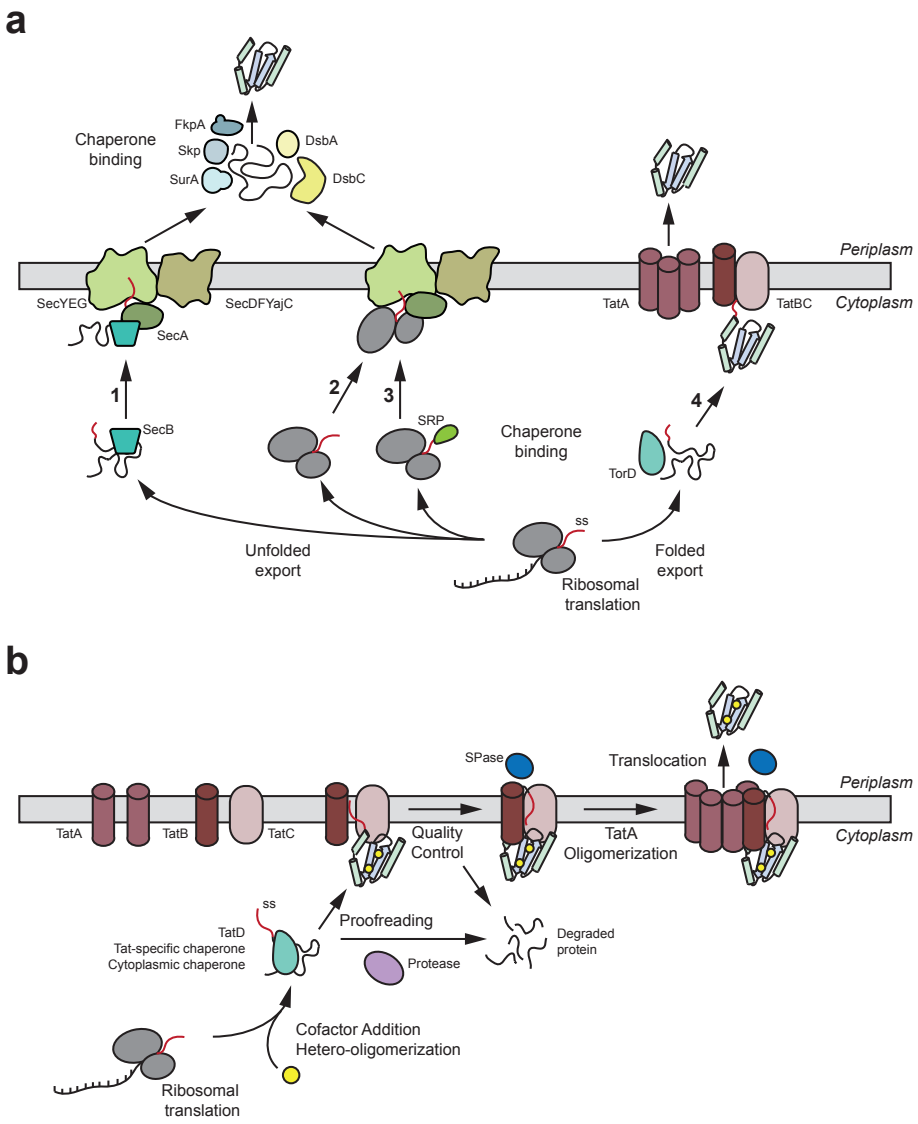

Figure 1.1: Biogenesis of periplasmic proteins in *E. coli*. A) Proteins destined for the periplasm are translated with N-terminal signal sequences (ss) (red line) that direct Sec, SRP-dependent, or Tat export. These signal peptides are later removed by a signal peptidase (SPase). (1,2) Sec export is a post-translational secretion mechanism that involves unfolded substrates. Some Sec substrates remain unfolded with assistance from the SecB chaperone (1) while others are exported in a SecB-independent fashion (2). Sec export is accomplished by the Sec translocase, which together with the SecA ATPase, ratchets unfolded Sec substrates into the periplasm through a narrow diameter pore formed by SecYEG. Once in the periplasm, molecular chaperones (e.g., FkpA, Skp, SurA) and enzymes of the disulfide bond formation pathway (e.g., DsbA, DsbC) promote the correct folding of newly translocated Sec substrates. (3) SRP-dependent export is a co-translocational mechanism whereby ribosome nascent chain complexes (RNC) are targeted to the membrane via the signal recognition particle (SRP) and its receptor FtsY. At the inner membrane, the RNC docks at the Sec translocase and the newly translated substrate is directly injected into the periplasm. (4) Tat export is a post-translational mechanism that involves completely folded substrates. Export is accomplished by the Tat translocase (TatABC) and the current model is shown in B). Tat substrates first bind their cytoplasmically-derived cofactors, metal ions and/or hetero-oligomerization partners in the presence of chaperones through a mechanism known as proofreading. Once the protein is transport-competent it is sent to the TatBC complex (typically four copies of each protein) where the signal sequence becomes membrane incorportated. A final quality control checkpoint innate to the translocon senses the folding state of the protein prior to export. TatA proteins oligomerize to permit the translocation of the protein across the inner membrane due to PMF. The SPase cleaves the signal peptide and the protein is released into the periplasmic space.

various natural proteins are exported by the Tat pathway including redox enzymes, periplasmic binding proteins, enzymes involved in cell membrane biogenesis, and virulence factors [21]. Folding proteins prior to export offers several advantages over secretion of the nascent polypeptide chain into the periplasm [21]. First, it allows for complex and cytoplasmic-derived cofactors to be added to Tat-targeted proteins [23]. Similarly, it dictates the incorporation of specific metal cofactors that are abundant in the cytoplasm; the protein folds around these metal ions so as to not have them out-competed by non-desired metals that are at higher concentration in the periplasm. Second, the Tat pathway permits translocation of hetero-oligomeric species, which often need cofactors to fold appropriately and must bind their partners to become transport competent [21]. Third, it allows for proteins to interact with cytoplasmic chaperones allowing for folding pathways not capable in the periplasm [20]. Lastly, secreted proteins with fast folding kinetics can be translocated through Tat and avoid having to be unfolded prior to export through Sec [26].

After translation, Tat targeted proteins interact with a variety of general and specific cytoplasmic chaperones that assist in folding and cofactor incorporation (**Figure 1.1b**) [18, 21]. General chaperones, such as GroEL, DnaK and SlyD, bind in a signal sequence independent manner [27]; however, some chaperones, such as TorD or DsmD, bind the signal sequence and are specific to each Tat substrate to ensure proper cofactor addition and/or oligomerization [21, 28, 29]. This step is known as proofreading since the chaperones serve as a checkpoint to sense the competency of the protein prior to export. The Tat substrates are then targeted to a hetero-oligomeric complex of

membrane-bound TatB and TatC [30-32]. TatC interacts with the signal sequence of the protein and is hypothesized to incorporate it into the membrane [33]. TatB has been shown to interact with proteins when complexed with TatC and is thought to block premature signal peptide cleavage [31, 33].

Formation of a TatBC complex bound to a signal peptide leads to the recruitment and homopolymerization of TatA [21, 32]. There remains much debate about the translocation event and whether the protein is sent through a pore created by the TatA proteins or is simply escorted through a weakened region of the inner membrane containing multiple TatA proteins [18, 21]; TatA recruitment is independent of substrate size suggesting the latter mechanism [31]. Interaction with TatA and the translocation event occurs due to an influx of almost 80,000 protons per exported protein, which is maintained by the proton motive force (PMF) between the inner and outer membrane [18]. A signal peptidase then cleaves off the Tat signal sequence releasing the protein to its final destination [33]. In addition to the proofreading checkpoint, there is also a debated quality control checkpoint inherent to the TatABC translocon itself [34-38]. This serves as a final sensor of folding and competency for export; however, this mechanism is still poorly understood.

Protein quality control

Cellular quality control refers to the ability of cells to discriminate between folding states of proteins to degrade proteins that have folded improperly and are thus nonfunctional to recycle cellular resources [39]. The term originates from eukaryotic

cells and the endoplasmic reticulum associated degradation pathways, where misfolded or unassembled proteins are deemed unfit and degraded by the ubiquitin proteasome pathway [40]. In both prokaryotes and eukaryotes, proteins encounter several quality control checkpoints during their synthesis, folding, and cellular transport [18, 19, 40, 41]. Quality control mechanisms impart a selective advantage to microbial cells by preventing the buildup of unfit proteins and allowing for the removal and recycling of dysfunctional proteins. Protein quality control goes beyond monitoring whether a protein is properly folded or not. It can govern the rate of folding to allow correct folding of protein domains and keep proteins in export compatible states [8, 39]. It can ensure a protein is folded in the correct cellular compartment as well as sense co-factor addition and disulfide bond formation to make sure the protein will function properly [20]. In the context of Tat pathway discussed above, both the proofreading step and inherent quality control element of the TatABC translocon would be considered cellular quality control.

In addition to quality control providing a selective advantage to cells, it is also a boon for protein engineers [6, 42]. The sequence space that can be explored by protein libraries can be astronomically large [43] and methods to focus libraries are sought to increase the likelihood of identifying proteins with enhanced fitness and function [44]. Producing protein library members *in vivo* subjects them to inherent quality control mechanisms, effectively removing non-folded, and thus non-functional members from consideration. Popular screening techniques such as phage or yeast-surface display [45] rely on quality control to eliminate misfolded proteins; however, these variants are

still present during the time consuming screening step where binding is assayed for. Protein folding reporters have sought to overcome this limitation by coupling the folded state of a protein to that of a fusion partner with a readout such as fluorescence or growth resistance [6]. These protein folding reporters serve as an indirect measure of heterologous protein amount and can be used for strain improvement as well as direct engineering of proteins for production enhancement. Popular folding reporters include green fluorescent protein (GFP) [46], dihydrofolate reductase [47], and chloramphenicol acetyl transferase (Cat) [48] for cytoplasmic expression and β-lactamase (Bla) [49] for periplasmic-localized proteins.

However, there are significant drawbacks to using indirect genetic folding reporters such as protein fusion artifacts, false-positives of protease cleaved chimeras or active reporters as part of inclusion bodies [50, 51]. A second generation of folding reporters have been developed that sandwich a protein of interest between two halves of a split reporter gene [52-54]. Such strategies eliminate the question of protease cleavage and only upon the protein of interest folding, will the reporter be reassembled and active [53]. These split reporters are even more prone to protein misfolding due to the destabilized reporter fragments and necessitate carefully designed linkers to ensure fragment reassembly. Although these are attractive quality-control-based engineering strategies and are effective for many proteins, using authentic quality control mechanisms that directly monitor the folded state of a protein may be preferred to ensure heterologous production compatibility, enable focused protein libraries, and permit strain engineering [55].

In the past, our laboratory has used the quality control mechanism of the Tat pathway to engineer proteins for expression in *E. coli* [55-57]. The Tat pathway offers numerous checkpoints and chaperone interactions to promote the folding of proteins. Most importantly, it only permits the translocation of fully folded proteins, eliminating aggregated proteins from consideration [55]. We have developed a selection strategy based on this quality control mechanism using a tripartite fusion of a natural Tat-substrate signal sequence, protein of interest, and Bla enzyme (**Figure 1.2**) [55, 58]. Only upon proper folding of the entire protein fusion is the protein permitted to be exported to the periplasm where the Bla is active. Any protein fusion that misfolds or aggregates remains in the cytoplasm and is eliminated by antibiotic selection. Such a strategy has been used to assess the production of several recombinant proteins as well as to increase the cytoplasmic expression of an antibody fragment [55, 56, 59]. However, the increase in cytoplasmic fitness is not necessarily accompanied by functional gain or results in loss of activity [56, 59], necessitating the inclusion of a secondary screen for activity. This strategy, based on authentic quality control, holds the promise of rapidly focusing protein libraries to contain only folded and functional members for consideration in the low-throughput screening step.

Cellulase engineering

Cellulases are rapidly becoming the highest produced proteins world wide due to their application in creating renewable liquid-based fuels from biomass sources [60]. Cellulases naturally function to break down plant matter to return nutrients to soil or into

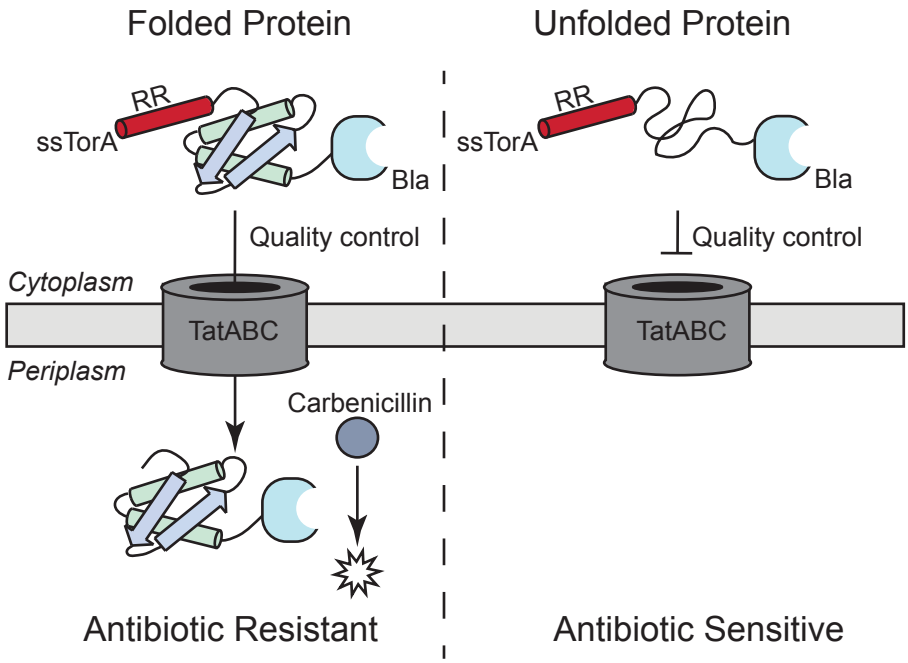

Figure 1.2: Tat-based genetic selction. This figure shows a schematic of a protein-folding reporter based on the authentic quality control mechanism of the Tat pathway. A gene encoding a protein of interest is sandwiched between that of the signal sequence for a natural Tat substrate TorA (ssTorA) and that of TEM1 beta-lactamase (Bla). The protein fusion is produced and folded in the cytoplasm prior to engaging the Tat translocon. Since the Tat pathway only exports folded protein, those that misfold or aggregate remain in the cytoplasm. Once exported to the periplasm, Bla acts on beta-lactam antibiotics (e.g. carbenicillin), enabling growth at specific selection conditions. Protein fusions that do not pass the quality control and are retained in the cytoplasm are sensitive to the selective amount of antibiotic, thus removing these members from future consideration.

digestible sugars [61]. Depolymerization of cellulose to glucose is thought to occur through the combination of enzymes from three classes (**Figure 1.3**) [62, 63]. First, an endocellulase (EC 3.2.1.4, also called an endo-β-1,4-glucanase) cleaves the bonds in amorphous regions of the cellulose fibril, forming new exposed ends that become the substrate sites for an exocellulase. Exocellulases (EC 3.2.1.91, also called an exo-β-1,4-glucanase or cellobiohydrolase) are specific for either the reducing or non-reducing end of the cellulose fibril, and deconstruct the cellulose into cellobiose molecules, which

are disaccharides of two glucose monomers. A third enzyme, a cellobiase (EC 3.2.1.21, also called a β-glucosidase), acts on the cellobiose and splits it into two glucose molecules. A mixture of these three enzymes results in the synergistic breakdown of cellulose to glucose to be used in yeast fermentation for the production of liquid fuels such as ethanol [60, 64]. Unfortunately, the enzymatic hydrolysis of biomass is not as simplistic as the aforementioned strategy [65] and the best known natural degraders of plant biomass produce an entire arsenal of plant cell wall degrading enzymes, including cellulases, xylanases, lignin peroxidases, and laccases [61, 66-68] to depolymerize recalcitrant cell walls as well as employ clever catalysis strategies such as cellulosomes to increase local concentrations of enzymes [69]. Additionally, freely-diffusing cellulases typically contain two domains, a catalytic domain which cleaves the bonds between sugars and a carbohydrate binding module (CBM) which localizes the cellulases near the cellulose fibril by binding directly to sugars (**Figure 1.3**) [60].

Creating optimal mixtures, often called cocktails, of cell wall degrading enzymes is crucial to the further development of biomass-based fuels to lower cost and increase yield [60]. Several strategies exist to improve cocktails, but foremost, each combination must be formulated for a specific biomass type and pretreatment strategy [62, 70]. Bioprospecting for cellulases and cellulose-degrading strategies has uncovered promising candidates to enhance breakdown of biomass [71, 72]. Heterologous expression of these enzymes in industrial, cellulase secreting organisms or conventional hosts has proven problematic and is a noted limitation in creating designer cocktails [16, 66, 69, 73, 74]. Additionally, non-enzymatic cellulases that promote degradation

through synergism and hitherto unknown catalytic mechanisms have also been discovered [65]. Protein engineers have sought to design cellulases with higher catalytic rates through directed evolution [75-77]; however, an incomplete mechanistic understanding of these enzymes and low-throughput screens on relevant substrates has stifled advancement [60, 70]. Further, these enzymes are being evolved to tolerate the processing conditions necessary for cellulose depolymerization including high temperature reactions and solvents such as ionic liquids [77]. Although a lot of attention and resources have been devoted to cellulase engineering, universal cocktails that are economically feasible for each type of biomass source are still currently limited [62].

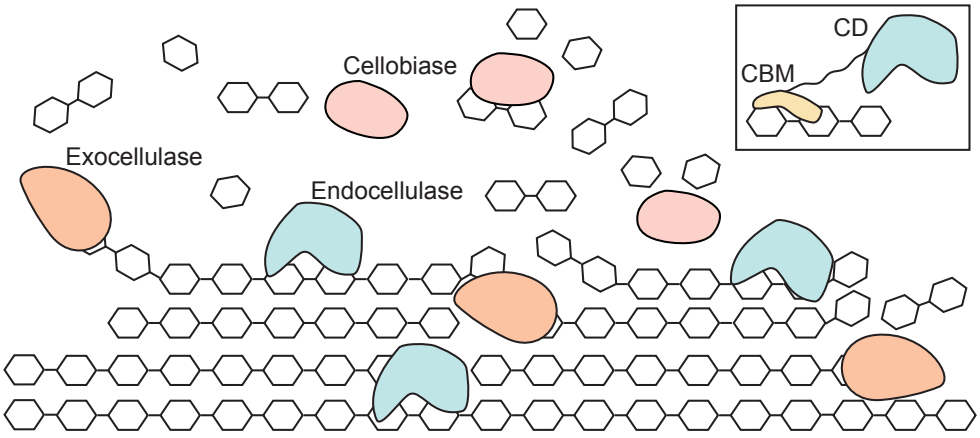

Figure 1.3: Schematic of cellulose degradation by cellulases. This figure depicts the degradation of a pure cellulose fibril by three classes of cellulases (adapted from a figure with permission by Jeremy Luterbacher). First an endocellulase (shown in blue) breaks bonds between cellulose fibrils and hydrolyzes bonds in the middle of the polymer to create new sugar ends. The exocellulase (shown in orange) then forms cellobiose molecules by hydrolyzing the fibril from the reducing or non-reducing end. An cellobiase (shown in coral) cleaves the cellobiose molecules resulting in two glucose monomers. All three classes of cellulases typically contain two domains, a CBM (shown in yellow) that attracts the cellulase to the cellulose fiber and a CD that performs the catalytic function.

CHAPTER 2

REPURPOSING BACTERIAL QUALITY CONTROL TO ENHANCE THE PRODUCTION OF A FUNGAL ENDOCELLULASE

Abstract

Plant pathogenic fungi are highly competent producers of lignocellulolytic enzymes that are required to breach the plant cell wall and surmount a successful invasion of their plant hosts. Unfortunately, the potential of these enzymes for industrial applications is undermined by low recombinant expression in heterologous hosts such as *E. coli*. To this end, we have created a simple and rapid two-tiered directed evolution strategy to enhance the production of a fungal endoclucanase from *Fusarium graminearum* in *E. coli.* The first tier involves a genetic selection for protein folding and stability that is based on the quality control mechanism of the twin-arginine translocation (Tat) pathway, while the second is a high-throughput screen for cellulase activity. Following two iterations through our selection and screen, we isolated a Cel5A variant whose production is increased 30-fold over the parent enzyme with just two amino acid substitutions. The gain in production is achieved without any compromise in activity on soluble or insoluble cellulose substrates, underscoring the value in this two-step evolution approach. Further, we have characterized the biophysical properties of the improved cellulases to begin determining which characteristics are selected by the Tat quality control mechanism to inform future protein engineering attempts that exploit this

pathway. The rapid increase in recombinant enzyme production through subtle mutations that do not affect activity, underscoring the value of using host quality control to engineer proteins in the environment they will be made in. *This chapter is a manuscript in preparation for submission to the journal Protein Science.*

Introduction

Protein engineering has proved an invaluable tool to biotechnology and the study of biology over the past 40 years after the advent of recombinant DNA technologies. Many of these efforts have focused around enhancement of catalysis, modification of binding specificity, increased thermodynamic stability, or creation of novel function [44]. However, one underutilized purpose of protein engineering is to increase the soluble production of heterologous protein products, which remains one of the bottlenecks in the bioproduct pipeline [5]. Proper protein expression is inherently host dependent, being determined by the presence of specific chaperones, cofactors and modifications, avoidance of protease degradation, maintenance of desired oligomerization states in the dense intracellular environment, and compatibility of protein-folding kinetics [8]. As such, innate quality control mechanisms function to sense the folded state of a protein and recycle improperly folded protein to conserve cellular resources. Combining natural quality control with protein engineering seeks to provide a method to enhance production by creating a protein product that is compatible in its host environment [42].

The majority of directed evolution approaches including phage or surface display indirectly utilize quality control since protein variants that are underexpressed due to

misfolding are effectively removed from final consideration [45]. Unfortunately, non-folded members still remain for the "high cost" activity screening step, which is the most challenging to implement, time consuming, and low throughput. The advent of protein-fusion folding reporters has sought to overcome this challenge by coupling the folding of a protein-of-interest with an activity readout [46, 48, 52, 54, 78]. Such methods are prone to false positives due to protein-fusion solubility enhancements, non-coupled folding, or active reporters in aggregated proteins [50, 51]. It is desired to have a labile approach to eliminate misfolded proteins that takes advantage of authentic quality control mechanisms to increase protein production and narrow the search through sequence space by focusing on folded members. Further, advantages in folding stability due to *in vivo* evolution are thought to impart enhanced thermodynamic and kinetic properties, providing additional benefit to such engineering strategies [53].

To this end, a genetic-based selection using the quality control mechanism of the twin-arginine translocation (Tat) pathway was created [55]. Specifically, the Tat pathway preferentially transports folded substrates across the inner membrane of *E. coli* with remarkable quality control [34, 35] that can provide selection pressure for protein folding and highly expressed proteins [55-57, 59]. However, improved protein folding stability independent of protein function as was observed with intracellular single-chain Fv antibodies [56] and heavy-chain Fv antibodies [59]. Thus, a secondary screening step for activity is required to isolate highly expressed and functional proteins. Such a two-tiered approach has been used in the past to enhance the production of a G-coupled protein receptor *in vivo* while maintaining activity and thermodynamic stability

[79]. Although successful, screening required a specific, membrane-soluble reagent and several rounds of directed evolution. A broadly applicable and *in vivo* selection-based method to enhance folding is highly desired. Enzymes, as a target for this two-tiered directed evolution approach, offer the advantage of having assayable activity and the compatibility of many semi-high throughput screens that already exist.

Cellulases have received attention over the past 50 years due to their eminent use in the conversion of lignocellulosic biomass into fuels [60]. Commercial cellulase cocktails produced from *Trichoderma spp.* have been the workhorse of the industry due to high production and catalytic activity [80-82]; however, due to the heterogeneity in biomass sources and severity of processing conditions, new cellulases must be discovered for efficient and cost effective conversion [83]. To this end, bioprospecting studies seek to identify cellulases that aid in the degradation of specific feedstocks, while recombinant systems are being developed to produce and engineer the selected cellulases for industrial purposes [70, 76, 77]. A potentially rich source of novel enzymes for this type of supplementation is highly virulent plant pathogenic fungi that produce a variety of cell wall degrading enzymes (CWDEs) during the infection process *in planta*, rendering them highly competent lignocellulolytic organisms [84-90].

One major producer of cellulolytic enzymes is the phytopathogen *Fusarium graminearum*, which was shown to be highly efficient at degrading a variety of lignocellullosic materials including switch grass [71, 88]. In fact, hydrolytic activity of *F. graminearum* preparations was higher overall than the activity observed with *T. reesei* RUT C-30 [71] and comparable to the commercial enzyme preparation Spezyme CP

(our unpublished observations). These results suggest that *F. graminearum* produces one or more highly active cellulolytic enzymes that could be incorporated synergistically into enzyme cocktails tailored for specific pretreated cellulosic materials. Here, four candidate CWDEs were identified in the genome of *F. graminearum* that were predicted to have a significant contribution to cell wall breakdown: FG03795, a GH5 family endo-β-1,4-glucanase; FG00571, a GH7 family Cel7A(CBHI)-type cellulose 1,4-β-cellobiosidase; FG03695, a GH61 family endo-1,4-β-cellulase type protein IV; and FG03003, a GH43 xylan 1,4-β-xylosidase. Unfortunately, a major impediment to the advancement of these and other related CWDEs is the fact that their expression in a tractable host such as *E. coli* is often met with low solubility, improper folding, and weak activity.

To remedy this problem, we developed a simple, two-tiered protein engineering strategy based on the Tat quality control mechanism for rapid laboratory evolution of cellulolytic enzyme variants. In light of the generally accepted principle of stability-function tradeoffs [91-93], our strategy seeks to increase protein stability and reduce its constraining effects in a manner that does not impair protein function and that may even facilitate the engineering of enhanced function [94]. By fusing a cellulolytic enzyme target to the N-terminus of mature TEM-1 β-lactamase (Bla) and using an N-terminal signal peptide to target the fusion to the Tat translocase, we show that it is possible to perform genetic selections for highly expressed enzymes. The second tier of our strategy is a semi-high-throughput functional screen using the soluble cellulase substrate carboxymethyl cellulose (CMC). By applying this second tier, our approach

favors the discovery of folding-enhanced soluble CWDEs in *E. coli* with uncompromised cellulolytic activity and bypasses previous problems with recombinant cellulase production in *E. coli* [77, 95-98]. Based on these results, this tool has the potential to provide a pipeline of redesigned CWDEs that have been optimized for heterologous expression in *E. coli* to be used in designer cellulase cocktails or as the starting point for further engineering efforts.

Results

Identification of candidate *F. graminearum* enzymes. Candidate genes from the *F. graminearum* genome were chosen based on: (i) their potential to contribute to cell wall digestion; (ii) direct analysis of this genus in comparison to other plant pathogenic genera; and (iii) a bioinformatics approach to identify the most promising candidates, focusing on those expressed *in planta* and by comparison to known enzymes using the CAzy (carbohydrate-active enzyme) database [99]. Our search identified four diverse enzyme candidates that appear to have a significant contribution to cell wall breakdown: FG03795; FG00571, FG03695; and FG03003. Of these, we focused on the glycoside hydrolase family 5 (GH5) enzyme FG03795 (hereafter called Cel5A) for the following reasons. First, Cel5A is highly expressed when the fungus is grown on cell walls and *in planta* during the infection process, indicating a strong natural selection with pathogenicity on grasses. Second, the GH5 family is abundant (at least 5 are expressed) in this fungus; GH5 is the most diverse GH family of cell wall degrading enzymes, including endocellulase EC 3.2.1.4, exocellulase EC 3.2.1.91, and

endoxylanase EC 3.2.1.8. Third, the GH5 family has the widely distributed $(\beta/\alpha)_8$ barrel structure, which has been proposed as a scaffold for both natural and artificial selection and evolution [100], and this fold is shared with enzymes in different GH families (GH1, 2, 5, 10, 26, 30, 35, 39, 42, 44, and 51, among others). Fourth, Cel5A shares sequence identity with PDB entries 1GZJ (catalytic domain of a GH5 from *Thermoascus aurantiacus*) and 1H1N (carbohydrate binding module (CBM) of a GH7 from *T. reesei*) enabling preliminary homology modeling of FG03795 (**Figure 2.1**).

Genetic selection of folding-enhanced Cel5A variants and activity screen. The first step of the two-tiered evolution strategy involves selecting Cel5A variants with increased stability using the Tat quality control mechanism (**Figure 2.1**). The wild-type *cel5A* gene from *F. graminearum* was codon optimized and cloned without its natural signal sequence as a fusion to the signal peptide of TorA, a natural Tat substrate, and to the *bla* gene to create a pSALect plasmid [55-57]. Using this reporter construct, we could rapidly test for production and folding by measuring growth sensitivity to the antibiotic carbenicillin. When the Cel5A wt protein fusion was expressed from pSALect, it grew at a relatively low carbenicillin concentration of 25 μg/mL and was sensitive to 50 μg/mL, which served as the selection pressure for our first round library (**Figure 2.2a**). Random mutagenesis was performed over the entire *cel5A* gene resulting in a library diversity of 1.7×10^6 members with a 0.3% DNA mutation rate.

The second step of the evolution strategy required a high-throughput screen to ensure the mutant Cel5As maintain their endocellulase activity (**Figure 2.1**). A 96-well

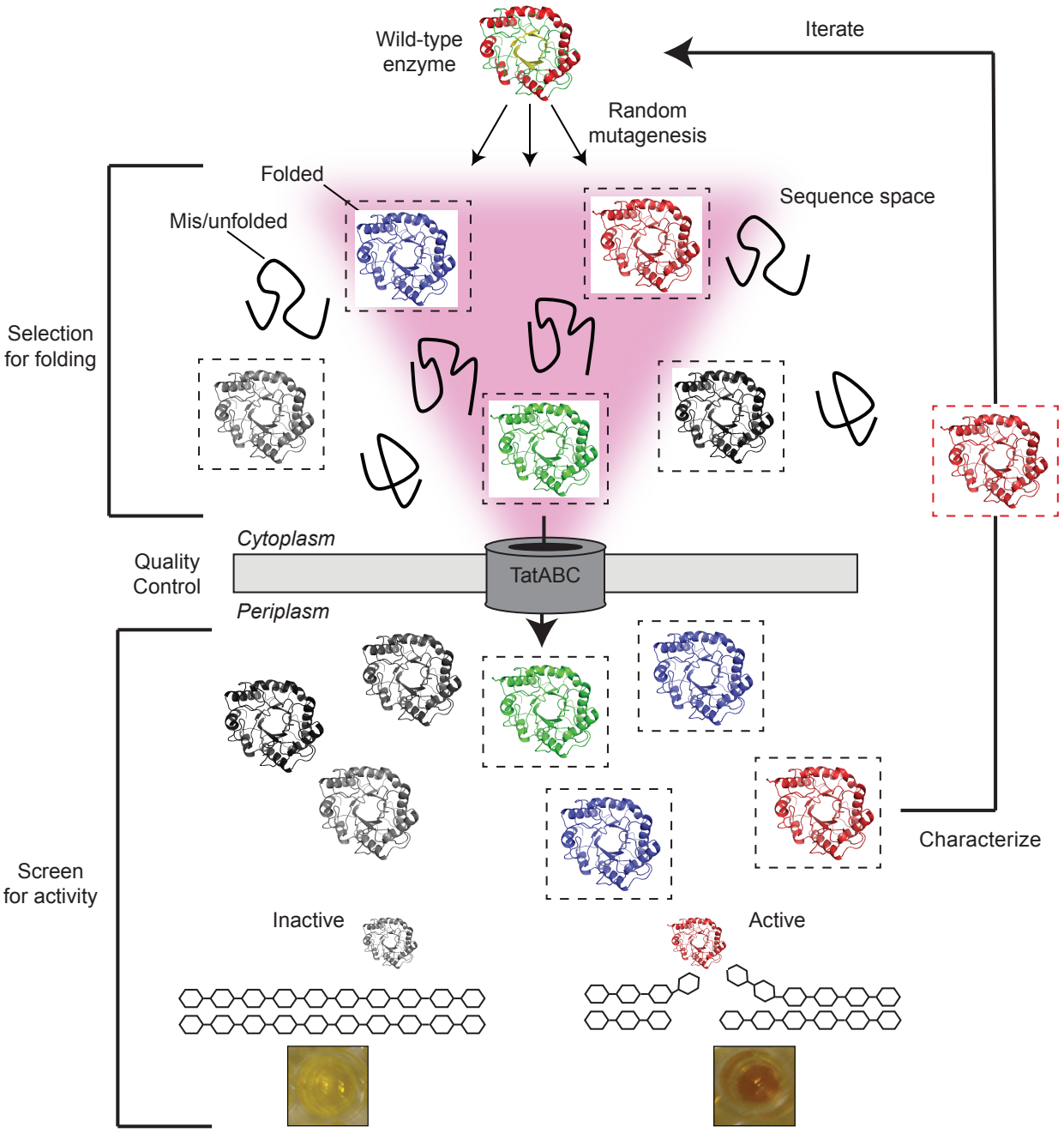

Figure 2.1: Schematic of the two-tiered directed evolution strategy. First, a Tat-based genetic selection is applied that links protein translocation with resistance to beta-lactam antibiotics. Since the quality control mechanism of the Tat pathway only permits the export of folded proteins, this genetic selection allows for the rapid and high-throughput isolation of well-folded, stable Cel5A library members while eliminating those that are poorly folded to focus the sequence space. A second screening step is imposed to ensure that the proteins that pass the Tat quality control retain high activity. For Cel5A, this involves a screen of enzyme activity using the soluble cellulose substrate CMC. Following isolation of active enzymes, the proteins are characterized for production enhancement and the best library member is used for further iterations through the two-step selection and screen. The protein structure depicted is the homology model for the catalytic domain of Cel5A.

plate-based screen using CMC followed by reaction with 3,5-dinitrosalicylic acid (DNS) was used to determine the amount of reducing ends produced by the cellulase [101]. From the selected library, 178 variants were grown separately and lysates were assayed for Cel5A activity (**Figure 2.2c**). Using this approach, members were isolated that maintained a similar level of activity compared to the parental enzyme. To confirm the growth phenotype was due to stability enhancement rather than strain mutation, we back transformed isolated pSALect plasmids for the most active variants and tested their sensitivity to antibiotic (**Figure 2.2a**)[57]. Using this criterion, we selected the top member from the first round library, C5A.1, which grew at the highest selective pressure (**Figure 2.2b)**. The C5A.1 protein contained two amino acid changes from the Cel5A wt, one in the CBM, G10D, and a second in the catalytic domain, T332I (**Table 2.1**).

C5A.1, when expressed from pSALect, offered an increase in antibiotic sensitivity to 75 μg/mL carbenicillin; however, this selective pressure is lower than many used in the past for other proteins expressed from pSALect [35, 56], suggesting a second round of evolution might further enhance the stability of the enzyme. Random mutagenesis was performed over the entire *c5A.1* gene resulting in a library size of 6.7×10^4 members with a similar error rate as the first round library. The library was selected at 100 μg/mL carbenicillin and 186 members were screened for retained cellulase activity. After back transformation, we isolated one variant, C5A.2, that had a substantial increase in antibiotic resistance beyond the selective pressure used to select the library and grew at high antibiotic concentrations observed for well expressed proteins (**Figure 2.2b**).

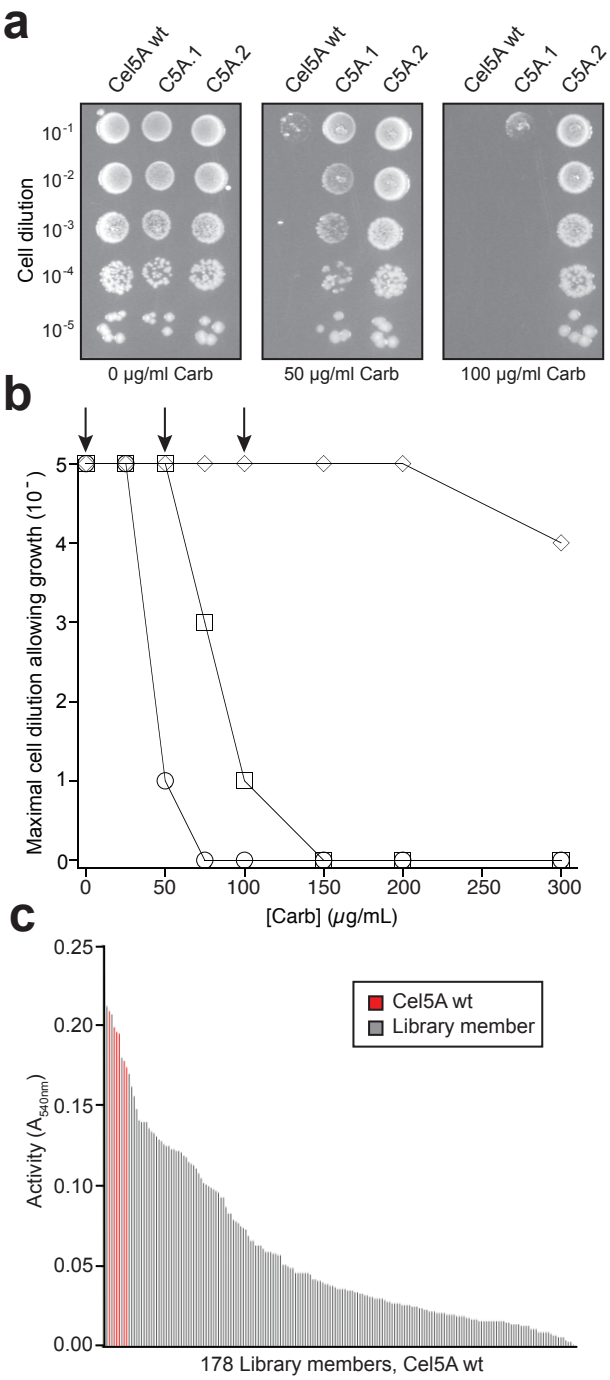

Figure 2.2: Quality control based selection library results. A) Spot plating results for Cel5A wt and the top hits for the first two libraries, C5A.1 and C5A.2, respectively. Normalized cultures of bacteria are serially diluted and plated on agar containing listed concentration of carbenicillin (Carb) antibiotic. The spot plates shown are for no selective antibiotic (0 µg/mL Carb) and the concentrations used for the two libraries (50 µg/mL Carb and 100 µg/mL Carb). B) Kill curves for Cel5A wt (open circle), C5A.1 (open square) and C5A.2 (open diamond). Plotted data represents the lowest cell dilution that allows for growth at each concentration of selective antibiotic and arrows above show location of spot plates above. C) Activity measured for library members selected from the Cel5A wt library. Grey bars are shown for library members and red bars for Cel5A wt. Activity is the absorbance at 540 nm after reaction with DNS and subtraction of a blank reaction.

C5A.2 contained an additional two amino acid changes from its parent C5A.1, both of which are in the catalytic domain, T166S and D342G (**Table 2.1**).

Table 2.1:

	Mutations		Production (U/L)	Specific Activity (U/nmol)	
	CBM	Catalytic Domain		50°C	37°C
Cel5A wt	-	-	0.15 ± 0.06	1.19 ± 0.12	0.61 ± 0.06
C5A.1	G10D	T332I	0.80 ± 0.44	1.28 ± 0.07	0.66 ± 0.04
C5A.2	G10D	T166S, T332I, D342G	4.49 ± 1.24	1.25 ± 0.06	0.63 ± 0.03
Cel5A TD	-	T332I, D342G	4.51 ± 0.11	1.21 ± 0.09	0.65 ± 0.05

Table 2.1: Cel5A mutations, production and specific activity. The table summarizes the mutations found in the proteins in this study, their cytoplasmic production as measured by activity on CMC and their specific activity on CMC. Mutations are shown as being part of the N-terminal CBM or C-terminal catalytic domain of Cel5A; no mutations were found in the linker between the two domains. Soluble cytoplasmic production of cellulase was determined through hydrolysis of CMC to a desired value and comparison of lysate volumes necessary to obtain such a value according to IUPAC standards. Hydrolysis was carried out using 1% CMC for 13 hours at 37°C and sugar reducing ends were determined by reaction with DNS. Error is the standard deviation from at least three biological replicated expression experiments. Specific activity on CMC was measured using purified enzymes at 50°C and 37°C using the IUPAC standard protocol for 1,4 β-endoglucanase on CMC. Error is the propagated error from three replicates of the experiment and error in the concentration measurement for the purified enzymes.

Enhanced production does not compromise cellulase activity on CMC. After two rounds of directed evolution, two engineered cellulases were isolated that showed greater antibiotic resistance using the Tat quality control-based folding reporter. To ensure production enhancement was not due to an increase in translocation or fusion to Bla, each Cel5A was expressed in the cytoplasm without the TorA-signal sequence or Bla reporter. The production of the Cel5A wt and isolated variants were determined by western blot analysis (**Figure 2.3a**). In the western blot, it is observed that there is a sequential increase in soluble production of the Cel5A through each round of evolution suggesting that the genetic selection worked as expected. When the insoluble fractions of these samples are compared, the lanes for the evolved proteins and wild-type are

identical indicating that the solubility of the enzyme has increased but not by partitioning out of the insoluble fraction. To quantify this increase in production, the amount of cellulase per culture volume was determined by assaying cell lysates on CMC (**Table 2.1**). Using volume of lysate required to catalyze a set amount of CMC, it was determined that the C5A.1 had almost a 5-fold increase in activity per culture volume compared to the Cel5A wt. Furthermore, the C5A.2 showed a similar 6-fold increase in lysate activity compared to its parent C5A.1.

In order to ensure that the increase in activity found above was not a result of altering the kinetics of the Cel5A during the evolution process, its specific activity was measured *in vitro*. The Cel5A wt and evolved variants were purified to greater that 85% purity using an N-terminally attached poly-histidine tag and affinity chromatography (**Figure 2.4**). By testing equal amounts of the Cel5A enzymes, it was possible to compare the specific activity of the three enzymes on CMC (**Table 2.1**). Endocellulase activity was characterized for reactions at both 50°C and 37°C, which are the standard reaction temperature for cellulases [102] and the temperature used for the activity screen, respectively. It was observed that all three cellulases have the same specific activity at both temperatures tested. This evidence suggests that the secondary screen merely ensured that activity was not lost after the genetic selection, but it did not enable find the most beneficial mutations (**Figure 2.5**). Analysis of the G10D and T332I mutation showed that neither completely recapitulated the cytoplasmic production increase found for C5A.1; however, the catalytic domain mutation of T332I did cause a

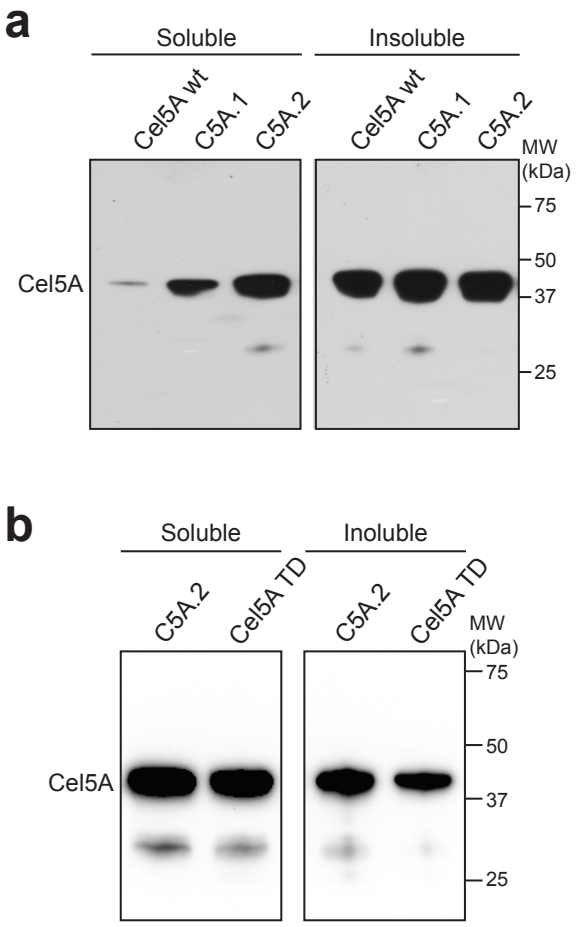

Figure 2.3: Soluble production enhancement of Cel5A. Western blots showing the soluble and insoluble production of Cel5A variants. Cel5A was produced in the cytoplasm lacking the Tat signal sequence or Bla-fusion protein used in the two-tiered selection and screen. Soluble cell lysates were normalized by total protein to ensure equal cell rupture. The insoluble fraction was loaded based on the total protein value for the soluble fraction. Each lane of the soluble fraction was loaded with 3 µg total soluble protein and the insoluble lanes contained 1 µg total soluble protein. Cel5A was detected using a C-terminal Flag epitope tag. A) Comparison of the production of Cel5A wt and isolated library members C5A.1 and C5A.2. B) Comparison of the production of the best evolved C5A.2 and Cel5A TD that contains only two amino acid substitutions and recapitulates the selection phenotype observed for C5A.2.

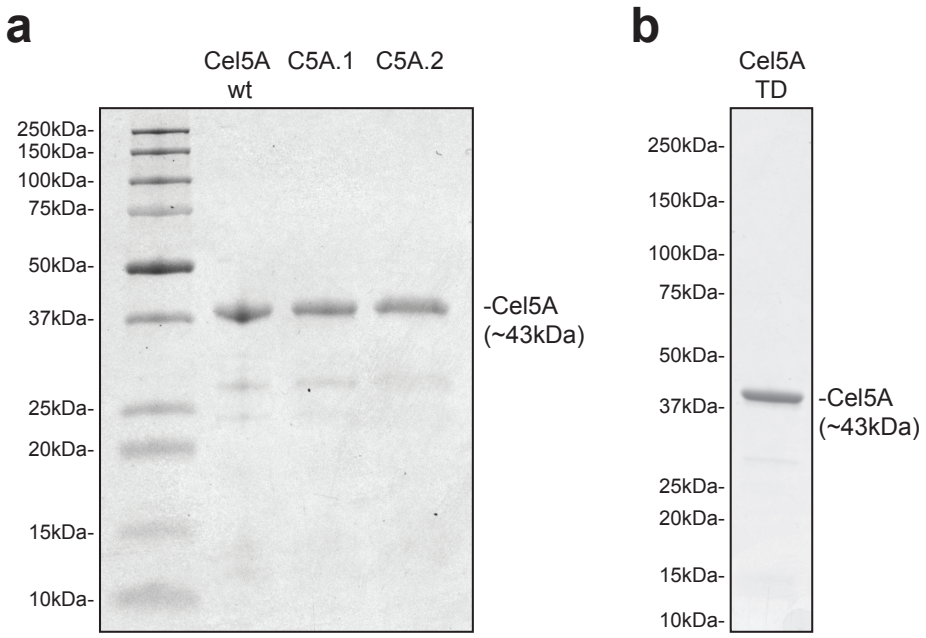

Figure 2.4: Purification of Cel5As. Cel5A enzymes were purified by IMAC, desalted into sodium acetate buffer and concentrated using a centrifugal molecular weight cutoff filter. Protein concentration was determined by BCA using a BSA standard. Protein purity was assessed by separating 3 µg of purified protein by SDS-PAGE and staining with Coomassie Blue. A) shows the relative purity and quantification of Cel5A wt, C5A.1, and C5A.2, while B) shows the purity of Cel5A TD.

the isolation of higher activity enzymes. Furthermore, since the specific activities of the three enzymes are comparable, the soluble lysate activity measurements (**Table 2.1**) suggests that through just two rounds of evolution the cytoplasmic production of Cel5A was enhanced by almost 30-fold.

After obtaining the C5A.2 protein with drastically increased production, it was desired to determine the cause of this enhancement. To find a direct cause it is imperative to only analyze mutations that give rise to the gain in production. Using site-directed mutagenesis it was possible to deconstruct the C5A.1 and C5A.2 proteins to

larger increase in production compared to D10G (**Figure 2.5**). The G10 is predicted be in a hinge region of the CBM and is thought to have steric clashes when mutated to the much larger aspartic acid. Additionally, this site is highly conserved as a glycine residue across *cel5A* homologs.

For all three of these reasons, the T332I mutation was focused on when deconstructing the C5A.2 variant. It was found that the addition of just the D342G mutation along with T332I could recapitulate the phenotypic growth on carbenicillin using pSALect as well as the cytoplasmic solubility observed for the best evolved variant C5A.2 (**Figure 2.5**). Further, when the production of the C5A.2 and Cel5A T332I D342G (Cel5A TD) were directly compared by activity on CMC (**Table 2.1**) or by western blot (**Figure 2.3b**), it was found that these were nearly identical. It is not surprising that the second round T166S substitution had little affect since it is a relatively conservative mutation. Cel5A TD was purified and had the same specific activity on CMC as the other Cel5As (**Table 2.1**). Thus, using the two-tiered selection and screening strategy we isolated a Cel5A variant containing only two required mutations that provided a 30-fold increase in soluble activity while not compromising activity on the substrate CMC.

Activity on CMC is often used to characterize endoglucanases due to it being soluble and catalyzed to a high degree by these enzymes; however, activity on CMC does not require CBM binding and is not representative of catalysis on insoluble substrates that are industrially relevant [60]. As such, the activity of the Cel5A wt and Cel5A TD was characterizedusing the insoluble substrates Avicel and bacterial

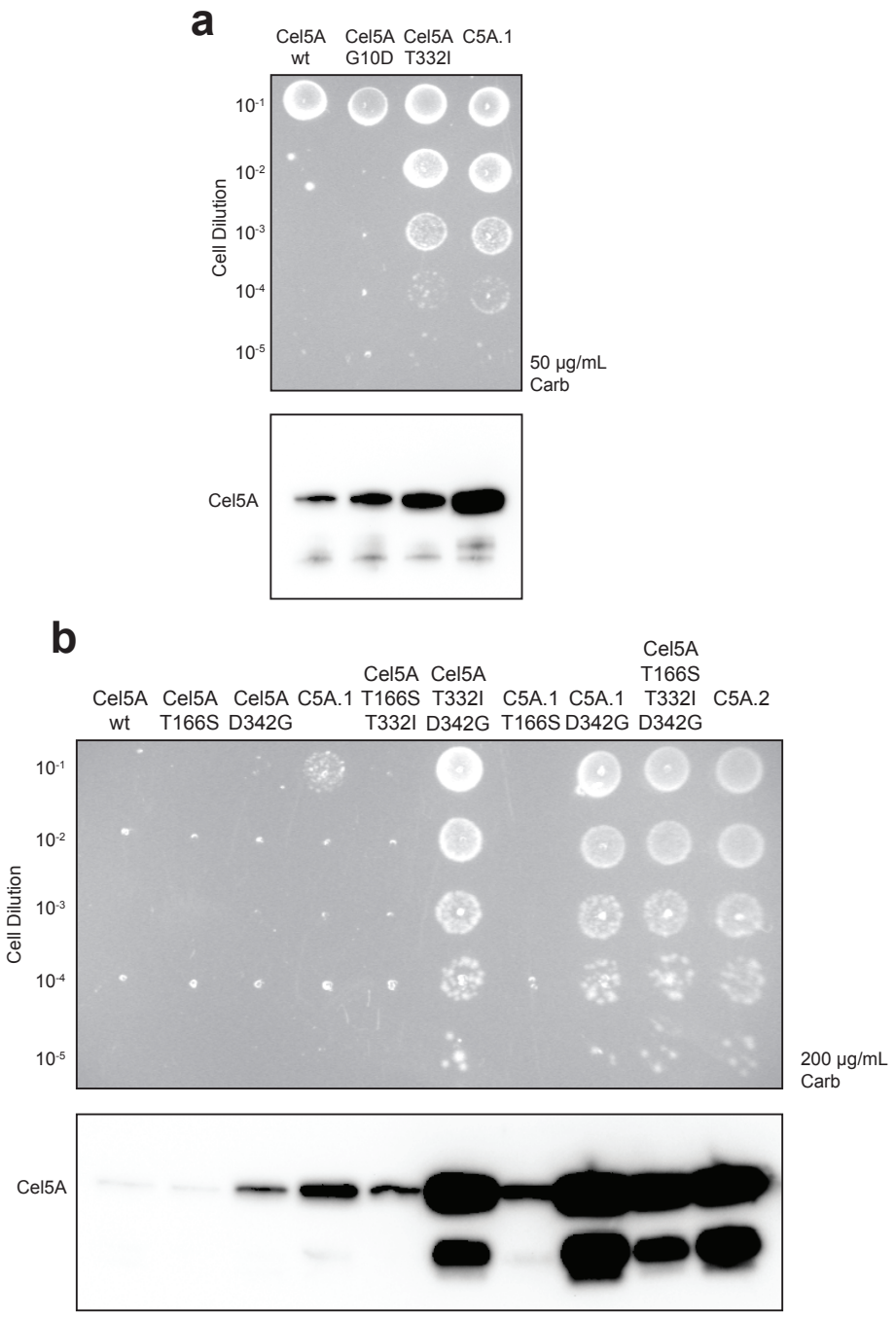

Figure 2.5: Deconstruction of evolved Cel5A variants. This figure shows the mutational analysis of A) C5A.1 and B) C5A.2 to determine which mutations are necessary for the observed increased antibiotic resistance using the pSALect system as well as enhanced soluble production. For spot plates, normalized cultures of bacteria are serially diluted and plated on agar containing listed concentration of carbenicillin antibiotic. Western blots show the cytoplasmic production of Cel5A without the Tat-signal sequence or Bla. Each lane contains 3 µg of total soluble protein, and the presence of Cel5A was detected using anti-Flag antibody. Since it was determined in A) that the T332I mutation best recapitulated the observed phenotypes for C5A.1, this mutation was focused on in the mutational analysis in B).

microcrystalline cellulose (BMCC) to ensure the activity of the enzyme was unchanged when tested using substrates not used for screening (**Figure 2.6**). As expected, the extent of reaction on Avicel and BMCC were low due to the endoglucanase being tested without an exoglucanase [64]. The extent of reaction was similar between the wild-type and best evolved Cel5A TD suggesting that the increase in folding stability did not have an impact on the ability of the cellulase to degrade insoluble substrates. Additionally, a beta-glucosidase was added for synergistic degradation of the substrates. It was found that the extent of reaction was increased through adding the additional enzyme and was not compromised by the mutations found (**Figure 2.6**).

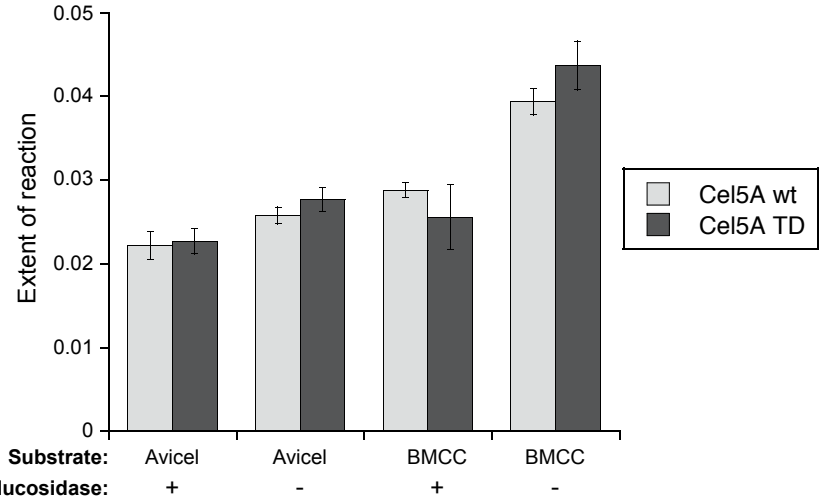

Figure 2.6: Extent of reaction on insoluble substrates. Results generated using the IUPAC standard protocol for insoluble substrates Avicel and BMCC using Cel5A wt (light grey) and Cel5A TD (dark grey). Protein loading was 340 nmol enzyme per gram Avicel substrate or 2,500 nmol enzyme per gram BMCC substrate. To demonstrate synergy, a beta-glucosidase was added to 30 CBU per gram substrate. Reducing sugars formed were detected using standard DNS protocol and the extent of reaction was calculated based on the theoretical maximum of reducing ends based on substrate loading. The error bars are the standard deviation from a minimum of three samples.

Biophysical characterization of evolved cellulases. The protein engineering strategy described above resulted in a rapid improvement in protein production while not sacrificing the activity of cellulase on soluble or insoluble substrates. Enhanced

production of a protein is a complex trait that can result from a combination of numerous factors such as thermodynamic stability, altered degradation or protein-protein interaction. Discovery of the underlying biophysical properties that govern enhanced enzyme production and a selected by the Tat-based reporter will hone future protein engineering attempts and shed light the poorly understood quality mechanism.

Minimization of folding free energy is the characteristic that drives the folding of proteins and further lowering this value may result in the increase in folding-enhanced proteins. A common way to test the thermodynamic stability of cellulases is to find a thermal inactivation temperature where half of their activity is lost (**Figure 2.7a**). Thermal stability of cellulases is an important design consideration for pretreatment strategies and for enzyme lifetime [77]. After incubation at higher temperatures, it was observed that the Cel5A wt had a thermal inactivation temperature of 68.1 ± 0.3°C, which is 4°C higher than that for the Cel5A TD (63.9 ± 0.1°C) (**Figure 2.7a**). Although, thermal stability is an important design principal, determination of the Gibb's free energy provides the folding energetics of the enzyme. The folding free energy was directly probed by measuring unfolding through intrinsic tryptophan fluorescence as a function of chemical denaturant (**Figure 2.7b**). It was found that the Cel5A wt unfolded at a denaturant concentration of 3.87 M guanidine HCl with a Gibb's free energy of 9.64 ± 0.64 kcal/mol. The evolved Cel5A TD denatured at a lower denaturant concentration of 2.8 M with a higher folding free energy of 10.01 ± 0.58 kcal/mol. From both of these measurements, it was found that the two catalytic domain mutations have caused the

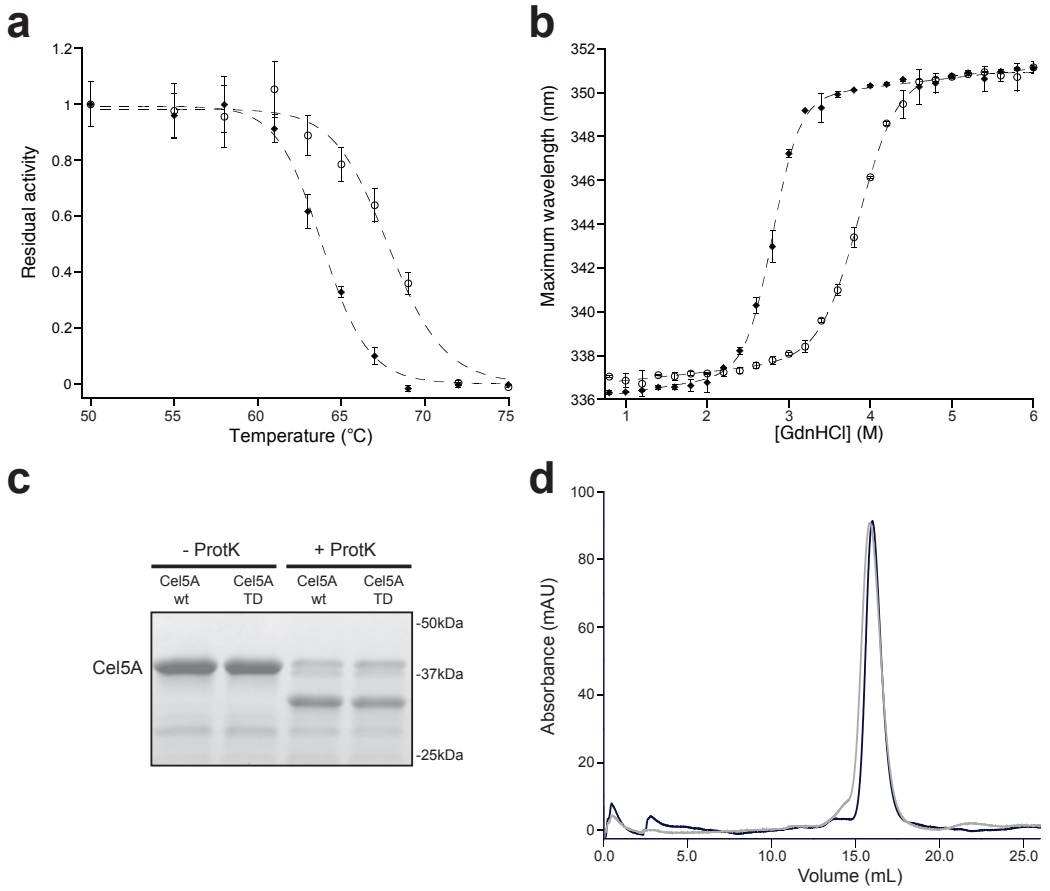

Figure 2.7: Biophysical characterization of Cel5A wt and Cel5A TD. A) Thermal stability of Cel5A wt (open circle) and Cel5A TD (closed diamond). Enzymes were incubated at indicated temperatures for 30 minutes prior to standard reaction with CMC. Residual activity is calculated by normalizing activity to that at 50°C. Error bars represent propagated error from the standard deviation of three samples. The data is fit using a Hill equation to calculate the TM where half of the residual activity remains. B) Thermodynamic stability of Cel5A wt (open circle) and Cel5A TD (closed diamond). Unfolding of the proteins was monitored by tryptophan (280 nm excitation) emission peak shift at increasing guanidine hydrochloride concentrations. Maximum peak emission was determined through curve fitting the emission spectra from 320 nm to 370 nm with a Taylor series expansion around the maximum wavelength. Error bars are the standard deviation of three samples and data was fit using a six-parameter thermodynamic model. C) Protease susceptibility of Cel5A wt and Cel5A TD. Samples were treated without (-ProtK) or with (+ProtK) ProteinaseK for 5 minutes at 37°C. Proteolytic degradation is shown by separating proteins by SDS-PAGE and staining the resulting gel with Coomassie Blue. D) Oligomerization of Cel5A wt (light grey) and Cel5A TD (dark grey). Resulting absorbance profile from separating 100 µg purified enzyme by size exclusion chromatography using a Superdex 75 column.

cellulase to lose overall thermodynamic stability in order to compensate for the increase in soluble production.

Another property that could cause the observed increase in folding stability of the enzymes would be a decrease in protease sensitivity, which was evaluated by treatment with Proteinase K (**Figure 2.7c**). It was found that the Cel5A wt and Cel5A TD were degraded to similar degrees suggesting non-specific protease sensitivity is not the cause for increased production. Both proteins showed measurable protease degradation to an approximately 30 kDa band, most likely showing the sensitivity of the less structured CBM and linker portions of the cellulase. Another possible reason for increased intracellular stability would be a decrease in aggregation of the proteins. Removal of proteins from higher oligomerization states could make them less susceptible to intracellular proteases as well as available for the selection strategy. The oligomerization state was evaluated by size exclusion chromatography using the Cel5A wt and Cel5A TD enzymes (**Figure 2.7d**). It was found that both proteins eluted similarly suggesting removal of aggregation is not an explanation for the production enhancement.

Genetic selection focuses protein library towards active members. In order to test how well the quality-control based folding reporter focuses protein libraries, the original Cel5A library was selected at several antibiotic concentrations and selected colonies were subjected to a plate-based secondary screen to determine the number of active members. One library was selected at a concentration of antibiotic at which the Cel5A

wt protein is not sensitive (25 μg/mL Carb) and was termed neutral selective pressure, since library members with equal or greater production to Cel5A wt would be found. Higher concentrations of antibiotic were also used to impose selective pressures above where Cel5A wt is sensitive. It was found that a moderate selective pressure of 50 μg/mL Carb reduced the number of selected colonies 4-fold compared to that at neutral selective pressure. A higher selective pressure of 100 μg/mL Carb decreased colonies 10-fold compared to neutral pressure. Selection for plasmid only was termed as no selective pressure and represents results as if the quality-control selection step was omitted.

When the percentage of active cellulases members from each library selection were compared, it was found that the neutral selective condition contained the most active members (**Figure 2.8**). Members that were not subjected to any selective pressure were 10% less likely to show any cellulase activity and 5% less likely to be as active as wild-type Cel5A when compared to those isolated at neutral selective pressure. This result suggests that selective pressure to retain production is tied with maintenance of enzyme activity. Moderate selective pressure was used in this work to isolate members whose production increased. Here it is found that moderate selective pressure resulted in less active members coming through the selection step, while high selective pressure caused a sharp decline in active members (**Figure 2.8**). Imposing too harsh a selection for folding enhancement seems to be compensated for by a loss of activity. Production increase through a large structural change could be found by rapidly changing selective pressure, but not surprisingly comes at the expense of

activity loss. These results suggest moderate gains in fitness through multiple rounds of directed evolution provide the best conditions for optimizing production while maintaining activity.

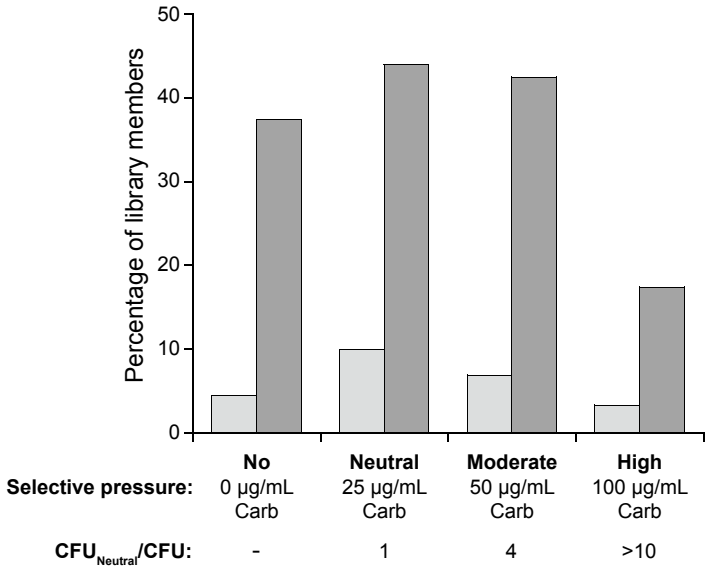

Figure 2.8: Library focusing using the Tat-based quality control selection. Percentage of library members that showed high activity (light grey bars) or weak activity (dark grey bars) for libraries selected at different antibiotic concentrations. Selections using the Cel5A wt random mutagenesis library were performed at no selective pressure (0 µg/mL Carb), neutral selective pressure (25 µg/mL Carb) (where the Cel5A naturally grows), moderate selective pressure (50 µg/mL Carb) (which was used to isolate C5A.1), and high selective pressure (100 µg/mL Carb). Activity was determined by clearance halos on CMC-containing LB agar. High activity is defined as having similar clearance halos to Cel5A wt on at least two of three plates. Weak activity is defined as showing some clearance on at least two of three plates. Library focusing is shown by calculating the ratio of colonies formed at the selective pressure compared to that at neutral selective pressure; a value could not be obtained for no selective pressure due to the plasmid encoded marker being used to provide resistance.

Discussion

The directed evolution approach described herein was used to rapidly enhance the soluble production of an endocellulase thirty-fold through just two rounds of directed evolution. Unlike previous studies that have utilized the Tat-quality control mechanism, this enhancement of production occurred without compromise of enzymatic activity on soluble or insoluble cellulose substrates due to the presence of a secondary activity screen. The increase in production occurred through two amino acid substitutions in the catalytic domain of Cel5A and was independent of the original genetic-based selection context. However, unlike previous work, the best-evolved variant showed a decrease in folding stability [53]. There seems to be a folding vs stability tradeoff during the evolution process [103]. It is hypothesized that folding free energy contributes to the ability of an enzyme to evolve new catalytic function [94], recognizing that a loss of stability often occurs when a binding pocket is remodeled [103]. Here, the discovered new function is extending the solubility of a hypothetical secreted protein to an intracellular context. Such a change occurred in a region significantly far from the active site to not alter its enzyme activity, but resulted in global destabilization to compensate for solubility enhancement. Changes in protein structure may allow for *E. coli* chaperone interaction to increase folded production, or could result in removal of kinetic-folding traps. Regardless, intracellular protein production is not correlated directly with higher thermodynamic stability for the protein studied in depth here. It is desired to continue to explore which biophysical characteristics are enriched for using the quality-

control based folding reporter to determine how general the changes are and to inform future engineering approaches.

Implementation of a secondary screen allowed for the engineering of enzyme production with retained high cellulase activity. However, screening on soluble substrates such as CMC maybe limiting when translating enzymes to cellulose conversion on pretreated biomass. Refinements could include relevant pretreated substrates, synergistic cellulases and more realistic reaction conditions [60, 70, 76, 82]. The advent of robotic screening platforms [104] aims to simultaneously optimize many of these parameters to create high efficiency enzyme cocktails. Such screens could be easily adapted to include the genetic selection for folding and would be highly beneficial for future lignocellulytic-enzyme engineering studies. Additional stringency in the secondary screening requirements could allow for enhancement of properties such as conversion, heat tolerance, low pH activity or synergism to improve the approach prescribed above [75]. Regardless, current cellulase engineering efforts towards activity or other enzyme enhancements will not be biotechnologically relevant if they result in poor host production [77], necessitating a multi-tiered approach as prescribed here.

Although adaption of the two-tiered strategy discussed here to include a second generation cellulase screen is a logical progression, the true power of quality control-based evolution would be to create valuable intracellular enzymes. Recombinant enzymes are prone to be instable in new hosts limiting potential chemistries. Additionally, most of the control of metabolic pathways has come in the form of transcriptional or translational regulation [105]. Engineering enzyme production in an

intracellular context [4] aims to solve both these issues through better production of low expressed proteins as well as the possibility of controlling the desired protein level by modulating selective pressures. Such control is highly desired to provide proper enzyme stoichiometry, eliminate bottlenecks, and yet not overburden the cell by gross overexpression. Using such an *in vivo* engineering approach, at least one study reports a 1,000-fold increase in metabolic production [106]. Combination of the quality control selection described above with rational protein design and established secondary screens could provide a new way of thinking about pathway development beyond altering activity and transcriptional regulation.

Materials and Methods

Bacterial strains and plasmids. *E. coli* strain DH5α was used for all cloning including library creation, MC4100 for genetic selection and spot plating experiments, BL21(DE3) for cytoplasmic expression analysis and SHuffle (T7 express) for purification. For plate-based selection experiments, the plasmid pSALect was used and is described in detail in [55, 57]. The *cel5A* gene from *Fusarium graminearum* (BROAD locus FG03795; GenBank locus NT_086532) was codon optimized and synthesized by Genescript. The *cel5A* was PCR amplified without its natural signal sequence and cloned between the *XbaI* and *SalI* sites in pSALect. An error prone DNA library of the entire *cel5A* gene was created with a theoretical DNA mutation rate of 0.15% according to the protocol [107]. Library size was determined by plating the transformed library on chloramphenicol (Cm) containing LB agar plates, and the error rates were confirmed by sequencing the DNA

of naïve library members. For cytoplasmic expression analysis, a pET28a plasmid was modified to contain an N-terminal 6x poly-histidine tag followed by Flag-epitope tag. Library hits were cloned between the *NdeI* and *SalI* sites of the resulting pET28a plasmid followed immediately by a stop codon.

Cel5A library selection and secondary CMC activity screen. MC4100 cells were transformed with *cel5A*-library containing pSALect. Dilutions of recovered cells were plated on LB agar containing the desired selective concentration of carbenicillin to ensure cells grew as single colonies [57]. Plates were incubated overnight at 30ºC. Individual colonies were picked into 96-well plates containing LB media supplemented with Cm antibiotic. The following day, a 96-deep-well culture plate containing 1 mL LB with Cm antibiotic was subcutured with 10 μL of the overnight culture. The deep-well cultures were grown for 2 hours at 37ºC and then transferred to 30ºC for 8 additional hours. Cultures were centrifuged to pellet cells and pellets were frozen at -20ºC overnight. To lyse cells, 50 μL of BugBuster was added to each well and the plate shook at room temperature for 30 minutes. The soluble cell fraction was recovered by centrifugation. The activity of each sample was determined by performing an assay using CMC, which contained 10 μL soluble lysate, 10 μL water and 20 μL of 2% CMC, and was carried out for 24 hours at 37ºC. Detection of catalysis was performed as described below. The LB-agar plate-based secondary screen for CMC hydrolysis of selected colonies was performed exactly as described in [108].

Soluble cellulase production. BL21(DE3) cells were freshly transformed with pET28a containing desired *cel5A* variants. An overnight culture of cells was subcultured to an OD_{600} of 0.05 in baffled flasks containing LB and kanamycin antibiotic. Cultures were grown to an OD_{600} of ~ 0.5 at 37°C. Protein production was induced for using 0.1 mM IPTG and was carried out at 30°C for 2 hours. Cells were harvested by centrifugation and lysed using BugBuster. The soluble fraction was recovered by centrifugation and total protein determined by a Bradford assay with a BSA standard. The insoluble pellet was washed twice with 50 mM Tris (pH 8), 1 mM EDTA followed by extraction of the insoluble sample in 1xPBS with 2% SDS at 100°C for 10 minutes. Protein samples were normalized by total soluble protein, separated by SDS-PAGE, transferred to a PVDF membrane, and detected using an anti-Flag primary antibody and appropriate secondary antibodies.

To determine the amount of soluble cellulase produced, CMC hydrolysis was measured. Cell culture was carried out exactly as described above. Cells were harvested by centrifugation and the resulting cell pellet was resuspended in 50 mM sodium acetate (pH 5) buffer. Cells were ruptured by sonication and the soluble fraction was recovered by centrifugation at 16,000xg for 30 minutes. CMC hyrdrolysis was carried out as described below for purified samples except the reaction was carried out for 13 hours at 37°C. BL21(DE3) cells that do not contain a plasmid were used as a blank for each amount of lysate tested.

Cel5A purification and quantification. SHuffle (T7 express) cells were freshly transformed with pET28a *cel5A*. An overnight culture of cells was subcultured to an OD_{600} of 0.05 in flasks containing 1 L of Terrific Broth and kanamycin. Cultures were grown to an OD_{600} of ~ 0.7 at 37°C. Protein production was induced for using 0.1 mM IPTG and was carried out at room temperature for 24 hours. Cells were pelleted by centrifugation and pellets were frozen at -80°C. Cell pellets were resuspended in 50 mM sodium phosphate (pH 7.4), 0.5 M NaCl and 20 mM imidazole and lysed using a cell homogenizer. Lysates were cleared of cell debris by centrifugation at 30,000xg for 30 minutes and filtered through a 0.22 μm filter. Protein purification was carried out using an FPLC by IMAC followed by desalting into 50 mM sodium acetate (pH 5) buffer. Proteins were further concentrated using 3,000 kDa molecular weight cutoff centrifuge columns made of polyethylsulfone. Protein concentration was determined by BCA assay using a BSA standard curve. Size exclusion chromatography was performed using an FPLC and Superdex 75 column. Protein purity was determined by separating the enzymes by SDS-PAGE and staining the resulting gel with Coomassie Blue.

Cel5A activity assays. Specific activity on CMC and extents of reaction were measured according to IUPAC standard protocols for endoglucanases [102]. For CMC hydrolysis, various amounts of Cel5A were incubated with 1% CMC at 50°C or 37°C in 50 mM sodium acetate (pH 5) buffer. After reaction, samples were mixed in a 1:3 ratio with DNS to detect the number of reducing ends formed during the reaction. The DNS reaction was carried out in a 100°C water bath for 5 minutes. For each set of samples

tested, glucose standards and appropriate blank reactions were included. DNS-reacted samples were cooled and diluted in water prior to measuring the A_{540nm}. Glucose standards were used to create a standard to convert A_{540nm} to concentration of glucose.

To determine the extent of reaction on insoluble substrates Avicel and BMCC, specific enzyme loadings per gram substrate were used. For Avicel reactions, between 340 to 133 nmol enzyme per gram Avicel were tested. Prior to setting up the BMCC reaction, the mass per volume of BMCC was measured by drying the substrate. For BMCC reactions, between 5,000 to 1,250 nmol enzyme per gram BMCC were tested. To analyze a simple synergy between cellulases, a β-glucosidase (Sigma) was added to 30 CBU/g. Reactions were carried out for 17.5 hours at 50°C using screw capped tubes and a rotating platform. The amount of reducing ends formed were measured using DNS as described above, and the extent of reaction calculated based on the theoretical maximum number of ends formed for each substrate loading.

Biophysical characterization of Cel5A. To measure the thermal inactivation of Cel5A, samples were incubated at temperatures between 50°C and 75°C for 30 minutes. The cellulase was cooled and CMC hydrolysis was carried out for 30 minutes at 50°C. To find the melting temperature for the cellulases, the thermal inactivation data was fit using a Hill equation. Cel5A stability was determined by monitoring its unfolding as a function of the denaturant guanidine HCl, between 0.8 - 6 M. Denaturation occurred for 1 hour at 25°C. Fluorescence emission spectra were recorded at 1 nm intervals between 320 and 370 nm after excitation at 280 nm using a scan rate of 10 nm per

second; slit width were set at ± 3 nm. Resulting emission spectra were fit using a third-order Taylor series expansion around the wavelength where maximal fluorescence was observed as shown in [109]. Fluorescence maxima were plotted as a function of denaturant and directly fit using a six-parameter fit described in [110] to determine the Gibb's free energy for each protein. All curve fitting was done using KaleidaGraph. To test the protease sensitivity of Cel5A, ProteinaseK (10 ng/μL) digestions occurred for 5 minutes at 37°C and reactions were quenched by the addition of PMSF (5 μM). Degradation was quantified by separating 5 μg total protein by SDS-PAGE followed by detection with Coomassie stain.

CHAPTER 3

RATIONAL DESIGN OF TAT QUALITY CONTROL SUPPRESSORS LEADS TO THE DISOCVERY OF THE CHAPERONE-LIKE ACTIVITY OF TATB

Abstract

The twin-arginine translocation (Tat) pathway transports folded protein from the cytoplasm to periplasm in gram-negative bacteria. Prior to the translocation event, several cellular checkpoints exist to ensure the competency of the protein for export, including what has been termed proofreading and quality control. The quality control step is inherent to the Tat translocon, and is debated in literature as to whether it can discriminate between the folded state of proteins and its existence. In this report, we describe for the first time the chaperone-like binding of the membrane-extrinsic C-terminal tail of TatB by showing its interaction with folded and unfolded citrate synthase *in vitro*. This interaction is not due to the presence of a Tat signal sequence, nor is citrate synthase a natural Tat substrate, showing the generality of TatB binding. Further, a homologous section of TatA does not interact with folded or unfolded citrate synthase, providing evidence for their distinctive roles in quality control and export. Quality control suppression mutations in TatB were analyzed to correlate alterations in TatB binding with export of misfolded protein. This work has provided direct evidence for the ability of TatB to recognize the folded state of proteins, potentially holding transport competent proteins at the membrane to prevent diffusion back to the

cytoplasm and targeting misfolded proteins to proteolytic degradation as a final transport check point. *This chapter is work in progress towards a manuscript.*

Introduction

The twin-arginine translocation (Tat) pathway is differentiated from other protein secretion pathways by its remarkable ability to translocate folded proteins across the cytoplasmic membrane in bacteria. Cytoplasmic folding allows for the incorporation of cytoplasmically-derived cofactors, binding of desired metals, hetero-oligomerization and interaction with cytoplasmic chaperones prior to translocation [21]. These simple reasons necessitate a pathway that begins the assembly and folds proteins in the cytoplasm prior to their export. As such, a series of checkpoints ensure a protein is transport competent prior to translocation to prevent the waste of cellular resources on proteins that will misfunction when properly localized [13, 21]. The first checkpoint is to confirm that desired cofactors are added to the protein through a step called proofreading [111]. This step is accompanied by chaperone binding to the Tat signal sequence and to the protein itself [28, 29, 112]. Chaperone interaction often assists in the conjugation of cofactors; only upon proper addition and folding can the chaperone be released to allow the protein to continue on to the Tat translocon [111]. The proofreading mechanism also ensures hetero-oligomerization of complexes before exposing the signal peptide to transport the multi-domain protein [21].

The second checkpoint is a proposed intrinsic quality control mechanism present in the TatABC translocon and functions as a final check of folding prior to export [32,

34]. Previously, it was shown that alkaline phosphatase (PhoA) could only become transport competent upon formation of disulfide bonds which enabled the proper folding of the protein [34]. Additionally, it has been found that unfolded PhoA can be targeted to the TatABC translocon *in vitro*, but there is no evidence for the transport of this protein [36]. Another study found short and unfolded non-natural Tat substrates maybe able to pass this critical quality control step and be exported; however, export required an overexpression of TatABC which is not indicative of how the natural system functions [37]. Translocation was only possible if the proteins did not have exposed hydrophobic residues, which is typically an indication of misfolded protein or a protein that is membrane bound [37]. This has led to the interesting hypothesis that what was previously termed quality control serves the function of holding up proteins with exposed hydrophobic regions for membrane incorporation [21]. It has also been proposed that TatB binding to folded proteins masks signal peptide cleavage to ensure translocation occurs before the protein is released from the membrane [33].

As such, the quality control mechanism of the Tat pathway remains an open question, but has been probed directly through random mutagenesis [35]. Using a panel of *de novo* designed peptides with varying degrees of aggregation and folding stability, it was possible to interrogate the quality control mechanism of the Tat pathway directly [35]. Specifically, the TatABC did not permit the export of the molten globular and aggregation prone α3B. Random mutagenesis was performed over *tatABC* and variants were found that allowed the translocation of this previously incompetent substrate and termed quality control suppressors. These TatABC variants were further

classified based on their ability to export unfolded PhoA, discovering mutations that permitted its translocation due to the hypothesized elimination of quality control [35]. This evidence furthers the thought that the Tat translocon itself plays a discriminating role in the transport of proteins even though there are numerous prior checkpoints that ensure the transport competency of proteins. Additionally, the work provides possible indication for transporting small, unstructured proteins.

The direct exploration of the Tat quality control identified a region of TatB and TatC that contained more mutations than other regions of the complex [113]. As such, we focused on TatB and its potential role in quality control. TatB is a membrane spanning protein with a single transmembrane helix followed by a C-terminal tail that resides in the cytoplasm [114, 115]. The membrane-spanning portion of the protein has been shown to be essential for functional Tat transport, while the C-terminal tail can be truncated without loss of translocation [116]. The C-terminal portion is composed of an amphipathic helix, similar to TatA, followed by a largely unstructured region [115, 116]. We suspect that the transmembrane portion of the protein does not recognize the folded-state of proteins to be exported, and plays more of a role in the assembly of the TatBC complex and interaction with TatA [30, 32, 33, 115, 117, 118]. If a portion of the TatB protein is to participate in quality control, it is likely to be the C-terminal tail, which also contained a higher than average amount of mutants identified in the quality control study [30, 31, 113, 115]. Crosslinking studies have found regions of both the transmembrane domain and amphipathic helix of TatB that bind to precursor proteins [119], possibly involved in signal sequence cleavage [33]. The membrane-extrinsic

domain of TatB is also known to homodimerize [114-116]. A version of TatB was created that lacked the N-terminal transmembrane portion of the protein (amino acids 1-21), $TatB_{\Delta 21}$. Additionally, removal of the membrane span of the protein allowed $TatB_{\Delta 21}$ to be produced in the soluble fraction of the cytoplasm and to be used in chaperone binding assays without adding detergents. A C-terminal polyhistidine tag was added to $TatB_{\Delta 21}$ to allow the protein to be purified by immobilized metal affinity chromatography (IMAC).

To begin directly testing the role of TatB in quality control, we studied its potential function as a chaperone. Molecular chaperones, in the most general sense, bind unfolded or folded proteins to assist in the formation of the final folding state [39, 120]. They can recognize unfolded proteins, block hydrophobic interactions, and fold a protein with energy input, as in the case of GroEL/GroES [121]. Additionally, several chaperones bind to folded proteins causing the rearrangement of secondary structure and refolding of proteins such as DsbC that isomerizes disulfide bonds [20, 122]. Molecular chaperones such as TorD and DmsD are specific to the Tat pathway, binding to protein TorA and DmsA, resprectively [28, 29]. However, many general cytoplasmic chaperones such as DnaK, GroEL and SlyD have been shown to interact with Tat substrates prior to their export [21, 27]. The Tat operon contains its own chaperone, known as TatD that additionally may play a role in intrinsic quality control [38, 123]. Previously, quality control in secretion machinery has been discovered in the general secretory pathway and binding partner SecA, which binds transport incompetent substrates to eliminate them from the secretion pool [124]. Here we wished to study a

similar role for TatB as well as begin deciphering how the quality control suppressor mutations might alter this chaperone ability. Although a variety of assays for chaperone-like activity exist, we decided to study binding to citrate synthase due to its easy implementation and is a well-established chaperone assay [120].

Results

TatB has chaperone-like activity. To determine if TatB has the hypothesized ability to interact with folded and unfolded proteins as part of the inherent quality control mechanism of the Tat pathway, standard chaperone assays were performed using the enzyme citrate synthase [120]. Citrate synthase is a central enzyme in the Kreb's cycle and has been used to analyze the chaperone qualities of proteins due to the ease of measuring its thermal aggregation or refolding from a denatured state through light scattering. Citrate synthase undergoes a structural change and aggregation at 43°C that is slightly above physiological temperatures [120]. Most chaperones remain folded at 43°C, which allows for the assaying of binding to folded citrate synthase and subsequent suppression of aggregation. Aggregation of citrate synthase results in increased turbidity in a solution and can be quantified spectrophotometrically. Conversely, citrate synthase can be chemically denatured allowing for interactions with unfolded protein to be monitored. Upon rapid dilution of unfolded citrate synthase, it refolds to a folding-pathway intermediate that is aggregation prone allowing light scattering to be used to detect chaperones that bind the unfolded citrate synthase and block the refolding transition [120]. Using these two methods, we can rapidly analyze

the ability of TatB to suppress the thermal aggregation of citrate synthase through interaction with the folded protein as well as its binding to unfolded citrate synthase to inhibit its refolding. Since citrate synthase is not a natural Tat substrate nor does it contain a Tat signal sequence, interaction with TatB will prove its general recognition of folded and/or unfolded proteins.

To determine if the TatB $_{\Delta 21}$ protein is capable of recognizing folded protein, its ability to block the thermal aggregation of citrate synthase was assayed for (**Figure 3.1a**). It was found that when a stoichiometric amount of TatB $_{\Delta 21}$ was added to citrate synthase at 43°C, a moderate gain in aggregation was observed by an increase in light scattering. However, when an equal amount of a known non-binding control (BSA) was added, aggregation was not suppressed and a steady rise in light scattering was measured which was 3-fold greater than that for TatB $_{\Delta 21}$. BSA was also tested in five-fold molar excess of the citrate synthase, and similarly did not block aggregation (data not shown). Prior to the addition of citrate synthase, TatB $_{\Delta 21}$ or BSA was incubated at 43°C with no detectable change in light scattering, suggesting the protein is stable at the elevated temperature. Taken together, these results suggest that the C-terminal portion of TatB shows the general chaperone-like ability to bind folded protein.

In addition to binding folded proteins, chaperones are often classified by their interaction with unfolded proteins, normally to promote the folding of nascent polypeptide chains. We tested the binding of TatB $_{\Delta 21}$ to unfolded citrate synthase by monitoring its ability to inhibit the refolding of the chemically denatured protein (**Figure 3.1b**). As was found for the thermal aggregation of citrate synthase, the addition of

stoichiometric amounts of TatB $_{\Delta 21}$ resulted in only a moderate gain in light scattering compared to large increase found with the non-binding BSA control. After an initial rise in light scattering, no further increase occurred during the reaction suggesting an initial diffusion or binding limitation prior to inhibition of citrate synthase refolding. Combining these two observations, we have shown for the first time that the C-terminal portion of TatB has the general chaperone activity by binding to folded and unfolded protein *in vitro*.

TatA does not show chaperone-like activity despite structural similarity to TatB. Although the two proteins have a different function in protein translocation, TatA and TatB share structural homology [116]. Both proteins have an N-terminal transmembrane helix followed by an amphipathic helix. However, chimeric proteins that combine the transmembrane helix from TatA with the C-terminal tail from TatB or vice versa show almost no protein export [116]. Due to their functional difference and inability to be substituted for each other, we suspected that TatA would not show chaperone-like binding. It has been suggested that TatA plays a passive role in transport, and quality control occurs prior to TatA oligomerization [30, 33]. To test this hypothesis, a version of TatA was expressed lacking its N-terminal membrane portion (amino acids 1-21), TatA$_{\Delta 21}$, and purified by IMAC. The ability of TatA $_{\Delta 21}$ to bind folded protein was assayed using the thermal aggregation of citrate synthase (**Figure 3.1c**). It was found that the TatA $_{\Delta 21}$ does not interact with folded citrate synthase as TatB $_{\Delta 21}$ does and behaves similar to the non-binding BSA control. This result supports the

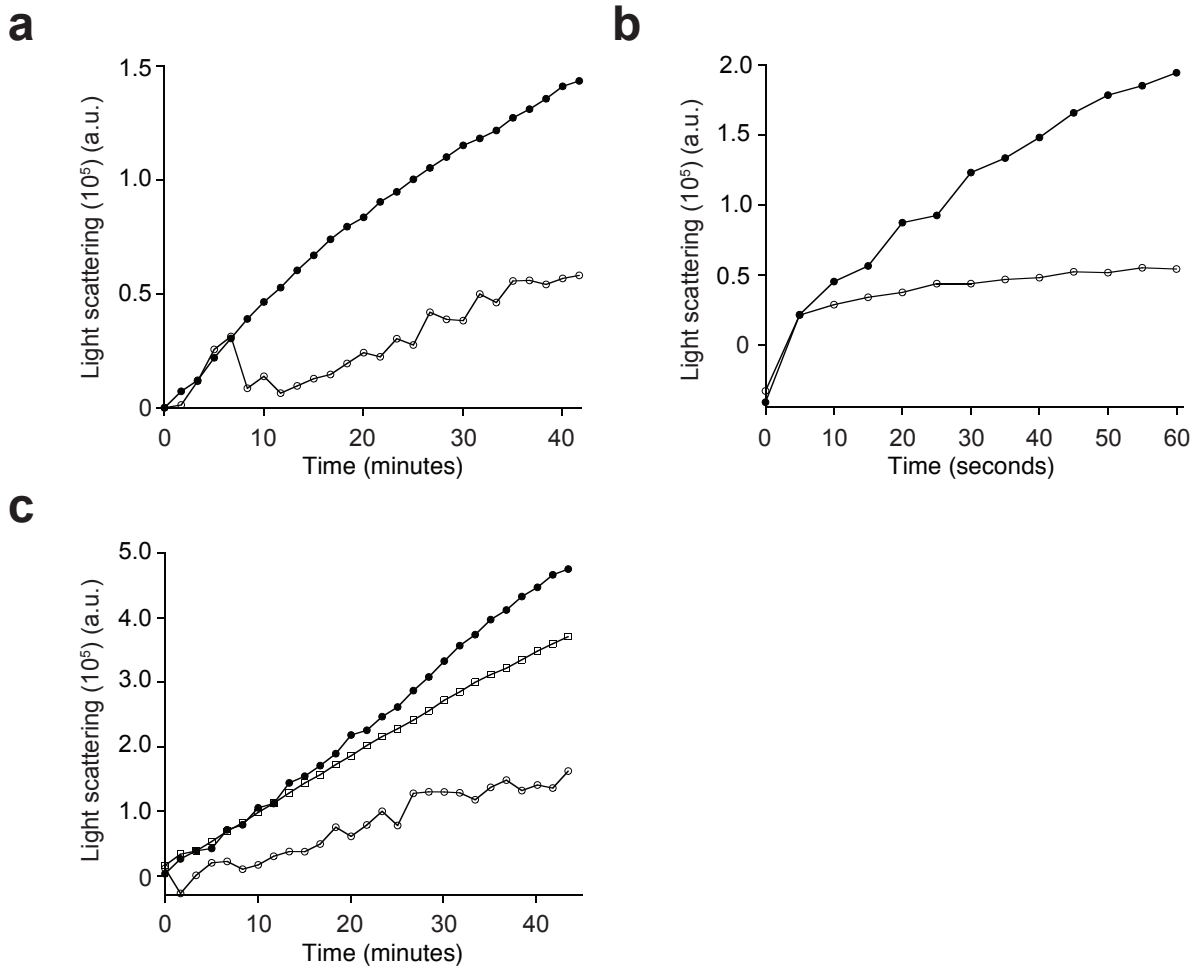

Figure 3.1: Chaperone-like activity of TatB. To analyze the chaperone-like activity of TatB to bind to folded or unfolded protein, assays using the enzyme citrate synthase (0.3 μM) were performed. Light scattering was measured as side scatter at 500 nm in arbitraty units (a.u.). A) Thermal aggregation of citrate synthase in the presence of 0.3 μM BSA (closed circles) or 0.3 μM $TatB_{\Delta 21}$ (open circles). B) Refolding of denatured citrate synthase in the presence of 0.3 μM BSA (closed circles) or 0.3 μM $TatB_{\Delta 21}$ (open circles). C) Thermal aggregation of citrate synthase in presence of 0.3 μM BSA (closed circles), 0.3 μM $TatB_{\Delta 21}$ (open circles), or 0.3 μM $TatA_{\Delta 21}$ (open squares). Light scattering is measured through excitation and emission at 500 nm (± 3 nm) and subtraction of light scattering before citrate synthase is added to the reaction. Data points shown are from moving average of collected light scattering where data points were set every 1.6 minutes and average over a 5 second interval around that point.

idea that part of the function of TatB is to recognize the folding state of proteins while TatA has a more passive role. Additionally, this suggests that although the proteins share structural homology, the C-terminal parts of the protein do have distinctive responsibilities that may have been overlooked in past studies.

Rational design of quality control suppressors. The proposed function of the Tat quality control mechanism is to block export of unfolded proteins from the cytoplasm to the periplasm [21]. Previously, Tat quality control suppressors, which permitted the translocation of otherwise unfit proteins, had been isolated from a random mutagenesis of the *tatABC* genes and were classified by their ability to translocate unfit proteins. [35]. Class I suppressors exported the entire panel of α3 peptides and misfolded PhoA better than wildtype cells, while class II suppressors were able to still discriminate α3A from the other peptides and did not translocate misfolded PhoA [35]. Additionally from that study, it was observed that an unstructured region of TatB, from residues 100 to 130, contained more quality control suppressor mutations than found elsewhere in TatABC as well as has three out seven histidines in the entire TatABC operon. One of these histidines had been substituted by asparagine in set of mutations for quality control suppressor Mut 57 that overall had the phenotype of a class I suppressor. The mutation of *tatB*(H109N) was made individually in pPRATABC and its suppressor phenotype characterized using the genetic selection based on the α3 series of peptides as performed previously [113] (**Figure 3.2a**). It was found that TatB(H109N) is a class II suppressor since it is still able to discriminate α3A from α3C/D, but not α3B.

In addition to assessing the suppressor phenotype, it is important to ensure that native-Tat substrate translocation is not eliminated through the mutation to TatB. Deficient Tat export is known to result in the accumulation of AmiA and AmiC in the cytoplasm of *E. coli*, preventing them from localizing properly to cleave N-acetylmuramic acid in the periplasmic peptidoglycan, and resulting in a chaining phenotype of cells [21], as shown by DADE cells (MC4100 Δ*tatABCDE*) (**Figure 3.2a**). Upon complementation of DADE cells with plasmid-encoded *tatABC* (pPRATABC), Tat export is restored and the chaining phenotype is eliminated as single cells are detected. When DADE cells are complemented with pPRATABC containing the single H109N substitution in TatB, single cells are also observed suggesting that the transport of AmiA and AmiC is not hindered by the quality control suppressor mutation.

Previously, it was hypothesized that the histidine residues present in TatB (positions H109, H112, and H123) play a role in quality control due to their abundance in a region previously shown to be highly mutated in isolated quality control suppressors. To this end, H123A was mutated in TatB individually and found to be a class II quality control suppressor [113]. Additionally, a variant of TatB was created that lacked all three histidines, TatB (H109A, H112A, H123A) [113]. The triple histidine knockout of TatB showed qualities of a class I quality control suppressor and could not discriminate between any of the α3 peptides. When DADE cells were complimented with pPRATABC TatB (H109A, H112A, H123A), cells grew as singlets, suggesting normal Tat translocation was not impacted by the replacement of the three histidines.

Lastly, TatB (H109A, H112A) was tested for it ability to discriminate between the α3 peptides and it was found to behave similar to a class II suppressor [113].

To reveal which histidine substitution or combination is necessary to suppress the distinction of the α3 peptides as observed by the TatB (H109A, H112A, H123A), all the additional histidine mutation sets were created (**Figure 3.2b**). TatB (H109A) behaved similarly to TatB (H109N) that was isolated as a quality control suppressor. The TatB (H112A) substitution was able to translocate the entire panel of α3 peptides suggesting this mutation is critical in the discrimination of folded and unfolded proteins. When this substitution was combined with H123A, a similar growth phenotype was obtained where all four α3 peptides were exported equally well at the selection pressure tested. When H109A was combined with H123A, quality control suppression remained the same as shown for the mutants singularly. When the H109A and H112A substitutions were made individually in pPRATABC, natural Tat export was not inhibited (**Figure 3.2c**). This result is not surprising given all three histidine mutations could be combined without detrimental effects on cell separation. Taken together, the H112A mutation in TatB seems to play the largest role in quality control inhibition, and rational substitution of any of the histidine residues in the short stretch of TatB results in some quality control suppression.

TatB quality control suppressors maintain the ability to bind unfolded protein. To begin determining what properties of the quality control suppressors are modified to give rise to deficient recognition of unfolded or aggregated proteins, the $TatB_{\Delta 21}$ (H109A,

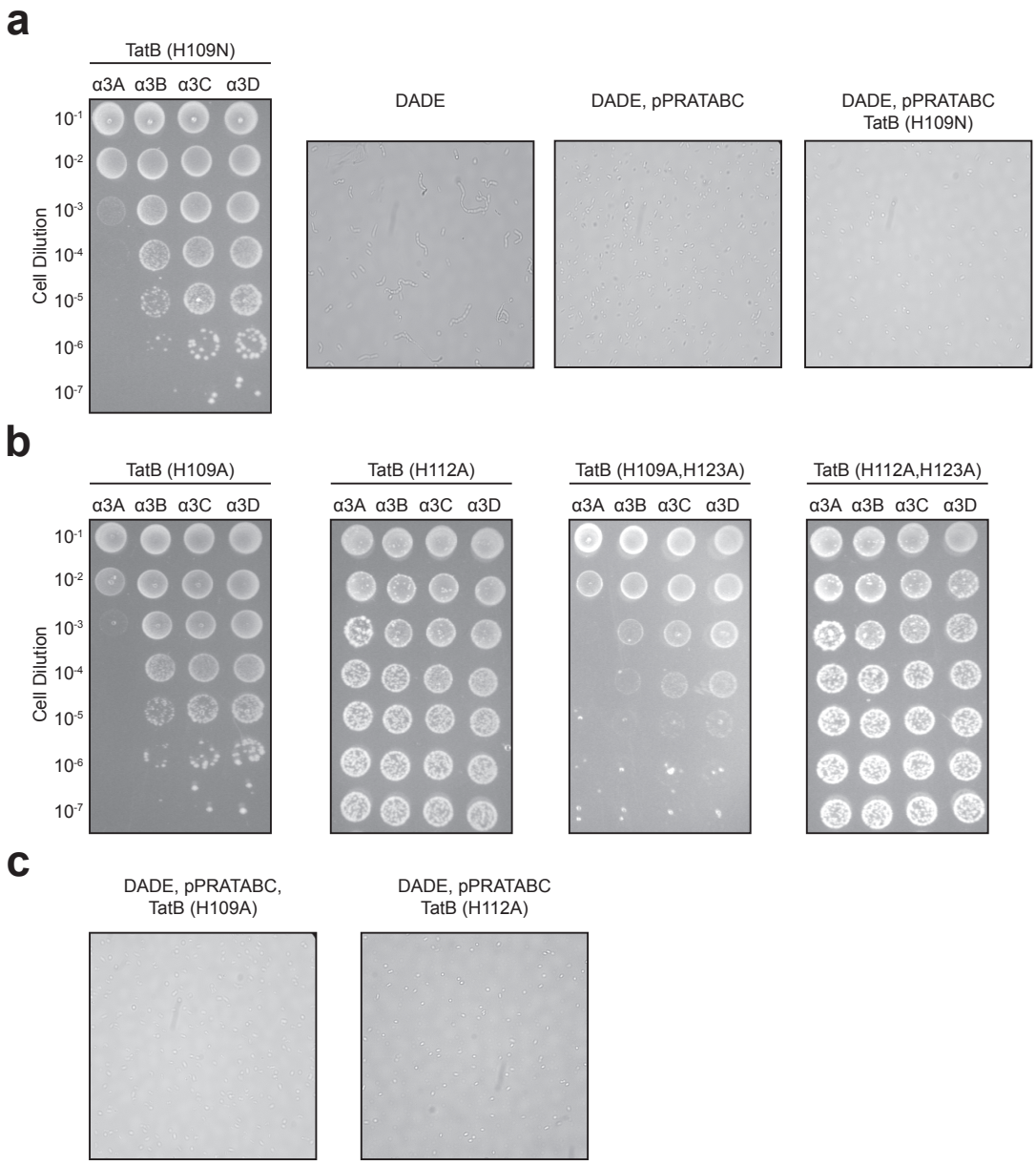

Figure 3.2: Rational design of Tat quality control suppressors. Quality control suppression was measured using ssTorA-α3-Bla reporter as described in [35]. All spot plates shown are at 300 μg/mL carbenicillin. Functional Tat complementation was analyzed by looking for the absence of chaining phenotype. A) Analysis of H109N mutation found in Mut 51 [35] for its role individually in quality control suppression. The DADE strain is ΔtatABCDE and can be complemented using the pPRATABC plasmid, which contains tatABC. Functional complementation of TatB (H109N) was preformed using the same pPRATABC plasmid containing a single point mutation and in DADE cells. B) Quality control suppression of histidine knockouts and combinations in TatB. C) Functional Tat complementation of single histidine knockouts in TatB.

H112A, H123A) protein was assayed for chaperone-like activity (**Figure 3.3a**). It was observed that the three-histidine mutations had a negative effect on TatB recognition of folded citrate synthase to inhibit its thermal aggregation. However, when binding of $TatB_{\Delta 21}$ (H109A, H112A, H123A) to chemically denatured citrate synthase was tested, it was found to suppress the refolding of citrate synthase similar to $TatB_{\Delta 21}$ (**Figure 3.3b**). This result suggests that the binding to unfolded protein is not perturbed by the triple histidine substitution. Additionally, $TatB_{\Delta 21}$ (H123A) was assayed in the citrate synthase refolding test and found to block the refolding of the enzyme. Surprisingly, for the quality control suppressors analyzed here, binding to unfolded protein was not altered while binding to folded citrate synthase was compromised.

$TatB_{\Delta 21}$ is highly aggregated. After $TatB_{\Delta 21}$ was purified by IMAC, the purity of the protein was assessed (**Figure 3.4a**). Surprisingly, when the protein was denatured, reduced, and separated by SDS-PAGE, it was observed that the protein did not migrate at the expected molecular mass of 17 kDa. Rather the protein seemed to behave similar to a trimer and had a measured mass of approximately 51 kDa. Additionally, a prominent band is found at the molecular mass of a dimer, 34 kDa. Previously, it has been found that the membrane-extrinsic domain of TatB binds to itself [114]; however, that interaction should be eliminated by boiling and SDS destabilization. A similar pattern was observed for purified $TatB_{\Delta 21}$ proteins that contained histidine mutations as well as for $TatA_{\Delta 21}$ (data not shown). TatB is known to oligomerize with TatC inside

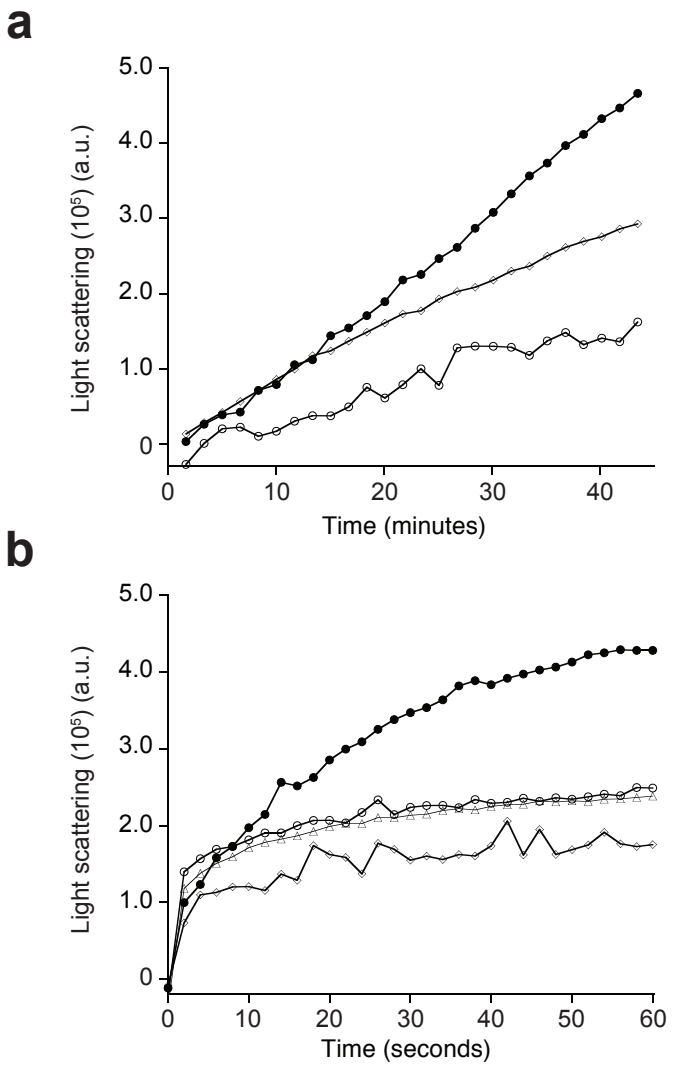

Figure 3.3: Chaperone-like activity of TatB quality control suppressors. A) Thermal aggregation of citrate synthase (0.3 μM) in the presence of 0.3 μM BSA (closed circles), 0.3 μM $TatB_{\Delta 21}$ (open circles), or 0.3 μM $TatB_{\Delta 21}$ (H109A, H112A, H123A) (open diamonds). B) Refolding of chemically denatured citrate synthase (0.3 μM) in the presence of 0.3 μM BSA (closed circles), 0.3 μM $TatB_{\Delta 21}$ (open circles), 0.3 μM $TatB_{\Delta 21}$ (H123A) (open triangles), or 0.3 μM $TatB_{\Delta 21}$ (H109A, H112A, H123A) (open diamonds).

cells in response to binding Tat signal sequences [30]; however, it was unexpected that the C-terminal tail would oligomerize and be denaturant resistant.

The purified $TatB_{\Delta 21}$ was separated by size exclusion chromatography to determine its oligomerization state in solution (**Figure 3.4b**). It was found that $TatB_{\Delta 21}$ eluted close to the void volume of the column suggesting it forms high molecular weight aggregates. Since it has been shown that $TatB_{\Delta 21}$ binds both folded and unfolded protein, it would be expected to oligomerize. Furthermore, when full length TatB or the membrane-extrinsic domain alone was studied previously *in vivo*, it was suggested that it homodimerized [30, 112, 115]; however, higher molecular weight oligomers could be possible as well. A small peak was detected at approximately 0.65 column volumes, which corresponds to molecular weight of 51 kDa that is similar to that observed by SDS-PAGE. Aggregation of protein maybe problematic in chaperone-activity assays since binding could be explained by interaction with non-folded protein. However, the C-terminal tail of TatB is not known to have much structure and when the aggregation of $TatB_{\Delta 21}$ was compared to $TatA_{\Delta 21}$, both proteins were found to aggregate similarly (data not shown). $TatA_{\Delta 21}$ did not show any binding to folded or unfolded citrate synthase (data not shown) suggesting that aggregation alone cannot explain the chaperone-like binding activity.

Genetic selection for factors that negatively affect Tat transport. Previously, our laboratory has focused on using the Tat pathway to increase the production of intracellular proteins [125] and isolate antibody fragments with enhanced properties using cleverly designed genetic selections [126] or screens [127]. Additionally, we have used these reporter proteins for mechanistic studies to find quality control suppressors

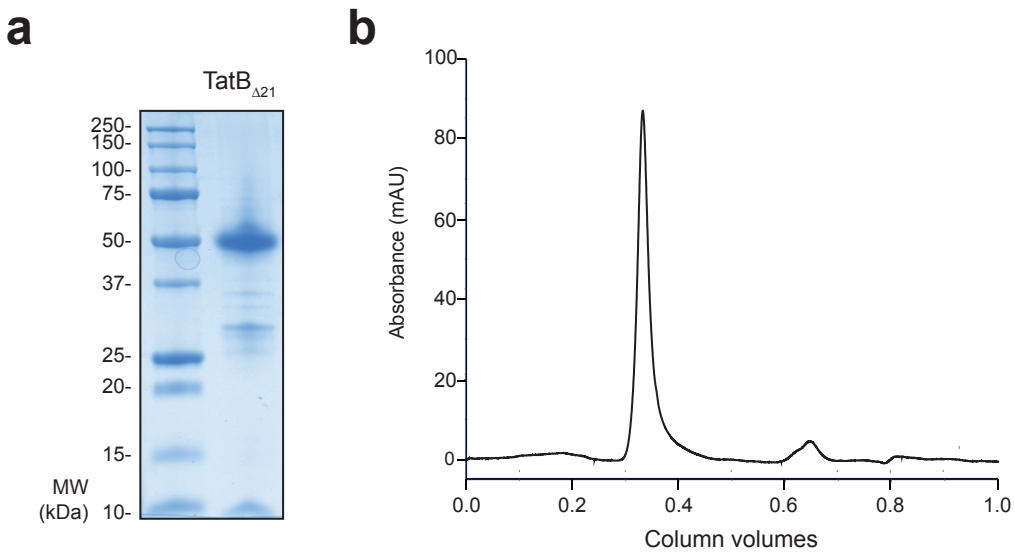

Figure 3.4: Purification of TatB$_{\Delta 21}$. A) Coomassie blue stained gel showing the purity of TatB$_{\Delta 21}$ (5 µg total protein) after IMAC. The expected molecular mass for TatB$_{\Delta 21}$ is approximately 17 kDa. B) Chromatogram resulting from size exclusion chromatography of purified TatB$_{\Delta 21}$. The void volume of the column is approximately 0.3 column volumes.

[35] and to discover Tat variants that hyper-export proteins [113]. Finding *cis* or *trans* factors that positively affect Tat translocation is biotechnologically desired [128], but more often pathway alterations will result in a loss of function. Because many questions still exist for the Tat mechanism of transport, including what endogenous *trans* factors assist in translocation of folded proteins besides specific chaperones [21], it is desired to have a genetic selection that reports the elimination of Tat transport [115, 129, 130]. Determining what proteins other the members of the Tat translocon can affect folding and export will allow those to be focused on for future engineering studies to enhance secretion. Additionally, such a reporter could be used to analyze a panel of proteins to find ones that do not pass the Tat quality control and remain in the cytoplasm.

A genetic reporter has been previously developed to analyze loss of Tat export using the enzyme chloramphenicol acetyl transferase (Cat) [129], which is a commonly used resistance marker in plasmids and hydrolyzes the antibiotic chloramphenicol (Cm). Cm inhibits protein synthesis by binding to the 50S segment of the ribosome and as such is a cytoplasmically acting antibiotic. A natural Tat substrate protein was fused to Cat, and only cytoplasmically-localized Cat confers Cm resistance [130]. This selection is in contrast to the Bla-based reporter previously utilized in our laboratory that hydrolyzes beta-lactam antibiotics only when it is found in the periplasm (**Figure 3.5a**) [129]. Cat has been fused to a full-length Tat protein to prove loss of function through Tat component knockout [129]. Additionally, Cat was fused directly to ssTorA to evaluate point mutants of TatB that result in inhibition of export [115].

Here we report the first use of the loss of Tat function genetic reporter for quality control analysis by fusing it to the α3 series of peptides. As a preliminary test of whether the strategy described could faithfully distinguish between functional and nonfunctional Tat export, the proposed Cat-based reporter was tested in *E. coli* with a fully functional Tat system (MC4100) and ones that lacked *tatC* (MC4100 Δ*tatC*), which have been shown to be transport deficient [22] (**Figure 3.5b-d**). In this experiment, we found the opposite resistance pattern that was originally observed for the α3 proteins, with α3A giving high resistance and the α3D giving lower resistance (**Figure 3.5b-c**). The resistance phenotype could occur due to an increase in translocation of the α3D compared to α3A, which has been found in the quality control studies. Another plausible explanation is α3D and α3C are produced at lower levels than that of α3A and

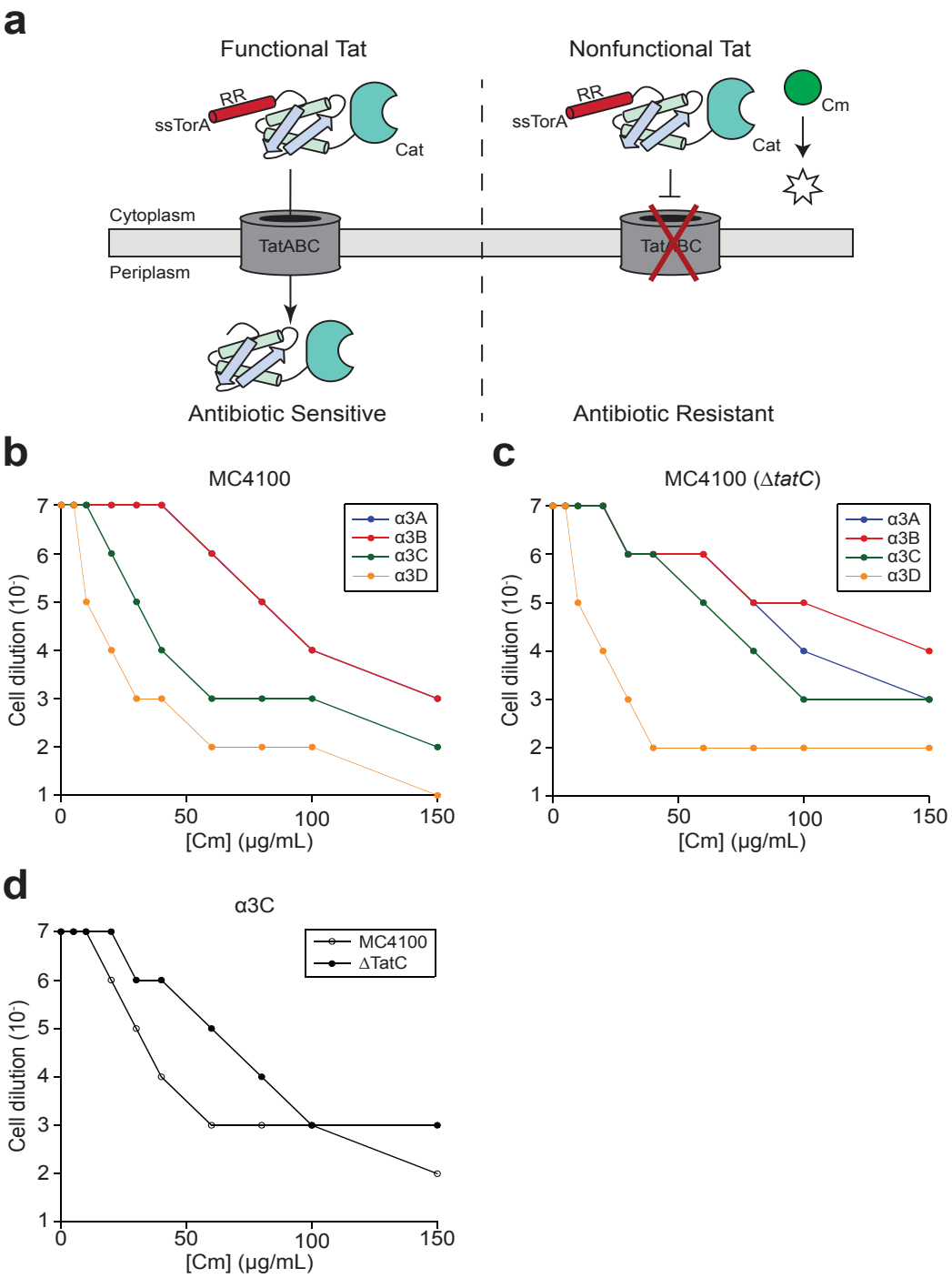

Figure 3.5: Genetic-based selection for loss of Tat export. A) Schematic of cytoplasmic-folding reporter used to assess functional Tat export. Using a similar strategy to the original Bla-based folding reporter, folded proteins are targeted to the Tat translocon using the signal sequence from TorA. The folding reporter is chloramphenicol acetyl transferase (Cat), which confers resistance to the antibiotic chloramphenicol (Cm). Cat is only active in the cytoplasm allowing for the observation of the opposite phenotype observed using the Bla-based folding strategy. Functional Tat transport will result in the Cat fusion accumulating in the periplasm and render the cells sensitive to Cm. Knockout of Tat export will leave the Cat reporter in the cytoplasm and cells will be resistant to Cm. B) Kill curves for the Cat-based folding reporter with the α3 series of proteins in cells with a functional Tat machinery (MC4100) and transport deficient cells (MC4100 *ΔtatB*). Data points shown are the maximal cell dilution that allowed for cell growth at each concentration of Cm tested. C) Comparison of ssTorA-α3C-Cat in MC4100 cells (open circles) and MC4100 *ΔtatB* cells (closed circles).

α3B (unpublished observations), giving rise to less reporter fusions and lower resistance. Most likely a combination of these affects result in the resistance profile observed. When the cells with functional Tat machinery are directly compared to those that are deficient, only when the α3C peptide is tested is a difference in growth observed (**Figure 3.5d**). As expected, cells with a complete Tat translocon are more sensitive to Cm antibiotic than those that cannot export protein. This result is preliminary evidence that a Cat-based reporter can successfully distinguish Tat-competent substrates as well as find negative effectors of Tat transport.

Discussion

In this work, we have shown for the first time the chaperone-like activity for the C-terminal portion of TatB by analyzing its binding to folded and unfolded citrate synthase. The chaperone-like ability of TatB is absent in TatA, even though the proteins are related, and further supports their distinct role in translocaton [32]. Chaperone-like activity is not dependent on the presence of a Tat signal sequence nor is citrate synthase a natural Tat substrate suggesting general binding of TatB. Protein-protein interaction provides a potential function for the C-terminus of TatB, which was previously proven not to be necessary for protein export and has been understudied [115, 116]. TatB has been shown to bind proteins when complexed with TatC [30, 31, 33, 112, 117, 118]. In addition to simply binding proteins, TatB-protein interactions are regulated since not all binding sites are simultaneously occupied in the TatBC complex

[31]. This has led to the hypothesis that TatB regulates export by retaining the protein in the cytoplasm [33, 131, 132].

It has been suggested that the Tat signal sequence orientates itself such that it can be processed by a signal peptidase prior to translocation [33]. If true, it would necessitate holding of the protein at the membrane to avoid diffusion back to the cytoplasm [33]. In fact, our laboratory has observed this phenomenon in what was termed a type I translocation intermediate [127]. When an epitope tag was placed directly after the Tat signal sequence, the tag could be detected on the periplasmic face of the membrane for a translocation-incompetent substrate; however, the tag could not be detected if it was moved to the C-terminus [127]. Additionally, detection of the type I translocation intermediate was not dependent on a functional Tat signal sequence suggesting membrane incorporation occurs prior to signal sequence recognition of TatC [127]. By binding folded protein at the membrane, TatB could prevent diffusion of protein with a cleaved signal sequence back to the cytoplasm prior to export. From the results gathered so far, we have put together a preliminary model of how TatB binding to folded protein could influence translocation (**Figure 3.6**). Although it is currently vague, it agrees with other evidence from Tat research and is the first showing the chaperone-like role of TatB in the quality control mechanism.

Furthermore, we have identified several rationally designed TatB quality control suppressors by targeting a region of that protein that was mutated with a high frequency in a past study [113]. Interestingly, when $TatB_{\Delta 21}$ (H109A, H112A, H123A) was assayed for chaperone activity, it was found that it no longer recognized folded citrate

synthase but retained the ability to inhibit the refolding of citrate synthase by binding to the unfolded protein. Quality control suppression is thought to arise due to an inability to recognize unfolded proteins; however, this property remains in-tact for the designed quality control suppressor. Further analysis of TatB (H109A, H112A, H123A) revealed that it is a class II suppressor since it does not result in increased amounts of transport-incompetent PhoA in the periplasm [113]. Class I quality control suppressors have both an increase of PhoA in the cytoplasm and periplasm compared to wild-type TatABC [35]. This observation suggests an interesting hypothesis that TatB may bind unfolded protein to prevent it from being translocated as well as induce the proteolytic degradation of the transport-incompetent substrate through a *cis* or *trans* mechanism (**Figure 3.6**). Class I suppressors are thought to lose this ability resulting in an abundance of protein that normally would be degraded, resulting in transport due to high cytoplasmic concentration [35]. To test this hypothesis, a class I quality control suppressor with mutations in TatB would need to be compared to current findings.

This suggested mechanism works with the established idea that the quality control mechanism recognizes exposed hydrophobic residues in unfolded proteins to occlude them from being translocated [37]. On the other hand, several studies have found TatB-independent degradation of misfolded proteins and suggest that the soluble chaperone TatD plays an active role in quality control [38, 123]. It has been proven that cytoplasmic chaperones play an active role in quality control prior to targeting the translocon, including degradation; herein, we suggest one final checkpoint of TatB binding to prevent export of misfolded protein. Lastly, it is suspected that protease

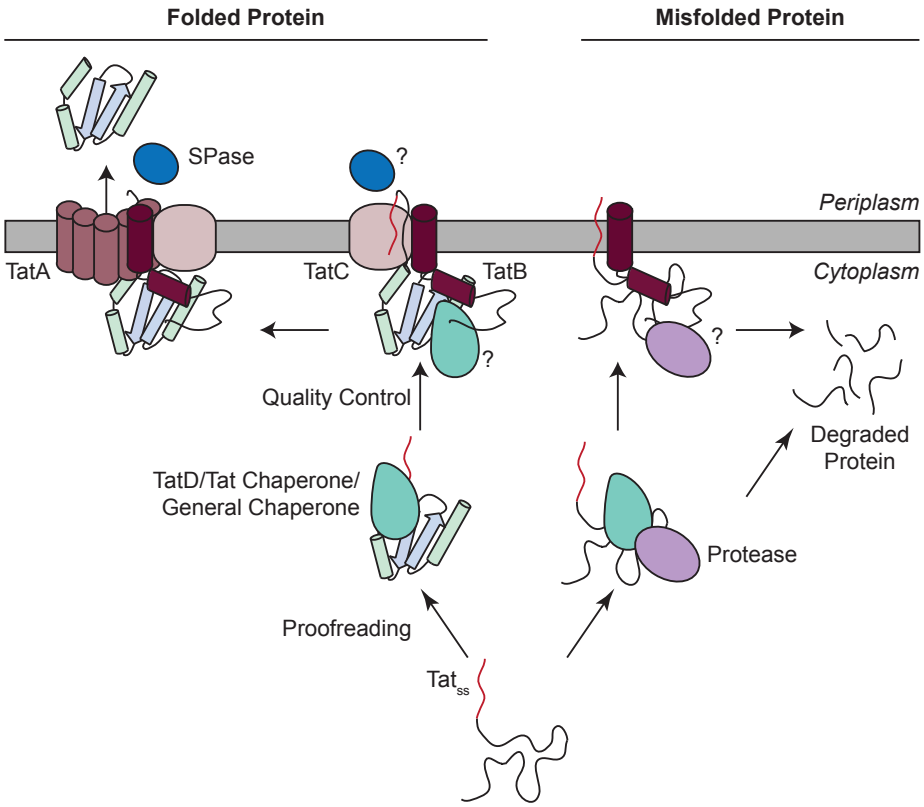

Figure 3.6: Predicted quality control in Tat pathway. This schematic combines the work in the chapter with others findings to show the role of TatB in quality control. Upon translation, a Tat substrate can follow two pathways, one where it folds properly and a second where it misfolds due to problems with one of the steps in the folded pathway. Folding, cofactor addition and oligomerization occur with the help of general-cytoplasmic and Tat-specific chaperones, which also monitor the completion of this process in a step known as proofreading. Then the protein is sent to the membrane where its signal sequence is membrane bound and directly interacts with TatC along with TatB. At this time TatB can interact with the folded protein, holding it in close proximity to the membrane. TatB binding to folded protein maybe desired if the signal sequence cleavage site is located in the periplasm to prevent premature exit of the protein. Additionally, TatB may bind chaperone proteins to facilitate passing the signal sequence to TatC. The TatBC complex causes an oligomerization of TatA and the subsequent translocation of protein from the cytoplasm to periplasm. In the event of improper folding or cofactor addition, the proofreading step may send the protein to be degraded. However, in the event unfolded protein reaches the membrane, TatB binds the unfolded protein and facilitates its degradation, possibly through interaction with a protease. Several questions still exist in this model including when the signal peptidase (SPase) cleaves the Tat-signal sequence, chaperone interaction with TatB, and proteolytic-degradation mechanism of unfolded substrates bound by TatB.

degradation maybe coupled to the Tat intrinsic quality control [113]; however, neither a specific nor general protease has been shown to interact with the Tat components.

Although we have shown direct binding of TatB to unfolded proteins, our hypothesis for this property does not completely explain the binding to folded protein and lack there-of in a type II quality control suppressor. As suggested earlier, this interaction with folded protein may hold the signal-peptidase-cleaved protein at the membrane prior to export. Additionally, evidence of TatB can recognizing chaperone proteins [112] suggests the possibility of transport-ready proteins being passed to TatB and/or regulation of export through chaperone interaction. Loss of binding to folded substrates would result in decreased retention at the membrane, but not necessarily an elimination of export [116] and it is unclear how this would promote α3B export.

We have begun probing TatB binding to proteins *in vivo* using a previously developed protein complementation assay [112] and the α3 peptides, but currently have no evidence for their direct binding to TatB inside cells (data not shown). Additionally, we have analyzed the ability of C-terminal truncations to export the α3D. We observed similar results to what was shown for the natural Tat substrate TorA [116], that export was lower in C-terminal deletions but was still possible (data not shown). Such findings support the hypothesis of holding substrates near the membrane and loss of this function resulting in less translocation. In the future we hope to probe TatB binding *in vitro* using techniques such as surface plasmon resonance to directly measure binding to many of the proteins discussed herein [29]. Using information from these studies, it will be possible to design *in vivo* experiments to further inform what proteins TatB

interacts with inside cells to give clues towards its role in translocation and quality control.

Although the results presented show the ability of TatB to interact with unfolded proteins, it does not provide an evolutionary explanation for this function. Tat-specific and general chaperones interact with Tat substrates prior to them engaging the translocon in a series of checkpoints to ensure the protein has the correct cofactor addition and/or is oligomerized properly prior to passing the protein to the TatBC complex [21]. Is the quality control recognition we are suggesting a final and redundant function? More likely, there are specific proteins or environmental conditions that might necessitate the presence of intrinsic quality control, as observed with PhoA [34]. Several Tat substrates do not have known chaperones suggesting TatB as the only checkpoint before the energy-intensive translocation step [13, 27, 111]. Additionally, this mechanism of binding unfolded protein could play a role in clearing the membrane in the event of faulty translocation or saturated translocons. Future studies should address the issue of what advantages additional quality control might have and how this is related to TatB interaction with unfolded protein [32, 36].

In addition to deciphering the function of the entire C-terminal portion of TatB, we have begun looking at the role of three histidines in binding to substrates and in quality control suppression. Several thoughts exist on the histidine abundance in TatB, including a function of binding divalent cations, a possible role as a serine protease [113], or even a signaling role through phosphorylation [133]. All of these possibilities will be further elucidated after the binding of folded and unfolded proteins to TatB is

completely deciphered. Lastly, we have created a genetic selection to isolate transport deficient substrates. This tool will prove valuable along with the Bla-based positive reporter to find *trans* factors such as proteases that play a critical role in the quality control mechanism of the Tat pathway.

Materials and methods

Bacterial strains and plasmids. MC4100 (Δ*tatABCDE*), referred to as DADE [134], was used for complementation studies and spot plating experiments. DADE(DE3) was used for expression of soluble Tat proteins from plasmids that contain a T7 promoter and to eliminate the possibility of co-purifying other Tat components. MC4100 (Δ*tatC*), referred to as B1LK0 [22], was used in Tat-export deficient studies. pPRATABC was previously described in [113]. pSALect α3 was created and described in [35]. Point mutations were introduced in *tatB* by QuikChange mutagenesis. *tatB* without its N-terminal membrane domain ($tatB_{\Delta 21}$) was amplified by PCR using the following oligonucleotide primers: 5'- gaggagcatatgccgcaacgactgcctgtgg and 5'-gaggagaagcttttagtgatgatgatgatgcggtttatcactcgacgaagggg. The second primer contained a C-terminal 6xHis tag and stop codon. The PCR product $tatB_{\Delta 21}$-*his* was added to pET22b using the NdeI and HindIII restriction enzyme sites. Similarly, $tatA_{\Delta 21}$-*his* was created and inserted into pET22b, except the following oligonucleotide primers were used: 5'- gaggagcatatgaccaaaaagctcggctccatc and 5'-gaggagaagcttttagtgatgatgatgatgcacctgctctttatcgtggcg.

Soluble Tat protein purification. DADE(DE3) cells were freshly transformed with pET22b that contained genes of the desired protein to be purified and were plated on LB agar containing ampicillin antibiotic (Amp). Individual colonies were selected and grown overnight in LB supplemented with Amp. Cells were subcultured to an OD_{600} of 0.05 in LB or terrific broth (TB) with Amp, and were incubated at 37°C for 2 hours. Protein expression was induced for using 0.1 mM IPTG and was carried out at 30°C for 6 hours. Cells were harvested at 5,000xg for 15 minutes due to DADE cells not pelleting as quickly as wildtype *E. coli*. Cell pellets were stored at -80°C. Cells were resuspended in in buffer containing 20 mM sodium phosphate, 0.5 M sodium chloride, 20 mM imidazole (pH7.4) and lysed using a homogenizer (Avestin). Soluble protein was recovered by centrifugation at 30,000xg for 30 minutes. Proteins were purified by IMAC using an AKTApurifier and HisTrap-HP columns (GE lifesciences) according to manufacturers' recommendations. $TatB_{\Delta 21}$ and $TatA_{\Delta 21}$ proteins were eluted in buffer containing 20 mM sodium phosphate, 0.5 M sodium chloride, 380 mM imidazole (pH7.4). Proteins were desalted into 40 mM HEPES-KOH (pH 7.4) using a HiPrep PD 26/10 desalting column (GE lifesciences) and concentrated using a 3,000 kDA molecular weight cutoff centrifuge filter (Millipore).

Citrate synthase (Sigma) was desalted into 50 mM Tris, 2 mM EDTA (pH 8.0) using a disposable PD-10 desalting column (GE lifesciences). The protein was concentrated using a 10,000 kDa molecular weight cutoff centrifuge filter (Sartorius). The protein was aliquoted and stored at -20°C [120]. Protein concentrations were determined by BCA assay(Sigma) using a BSA standard curve. Protein purity was

determined by separating proteins by SDS-PAGE and staining the resulting gel with Coomassie blue.

Citrate synthase chaperone assays. To analyze the influence of chaperones on the thermal aggregation of citrate synthase the following protocol was used. Citrate synthase was diluted to 30 μM in a buffer of 50 mM Tris, 2 mM EDTA (pH8) on ice. A solution of 40 mM HEPES-KOH (pH 7.5) was warmed to 43°C prior to addition of chaperone protein or BSA. The final concentration of the chaperone proteins was 0.3 μM and these were incubated for 5 minutes at 43°C. Light scattering was measured using a spectrofluorometer (PTI) with an excitation and emission of 500 nm (± 2nm). Prior to addition of citrate synthase, the aggregation of the chaperone proteins was monitored, which served as the blank for the reaction. The citrate synthase was diluted 1:200 into the chaperone-containing solution at 43°C and the light scattering was measured for approximately 40 minutes taking 2 to 5 data points per second.

To analyze the influence of chaperones on the refolding of chemically denatured citrate synthase the following protocol was used. Citrate synthase was denatured in 6 M guandidine HCl with 20 mM DTT for 4 hours on ice. A solution of 40 mM HEPES-KOH (pH 7.5) was warmed to 25°C prior to addition of chaperone protein or BSA. The final concentration of the chaperone proteins was 0.3 μM and these were incubated for 5 minutes at 25°C prior to the addition of CS. Fluorometer settings were the same as above, and data points were taken at a rate of 5 per second. The citrate synthase and 6 M denaturant was diluted 1:200 to a final concentration 0f 0.3 μM and 30 mM,

respectively. Light scattering was measured for approximately 3 minutes.

Both sets of data were analyzed using a MATLAB script that calculates a moving average at a specified frequency and interval to eliminate light scattering peaks that occurred randomly throughout the experiment.

Microscopy and spot plating. DADE cells were freshly transformed with pPRATABC containing desired *tatB* point mutations and plated on LB agar containing anhydrous tetracycline (aTc) antibiotic. Individual colonies were selected and grown overnight at 37°C in LB with aTc. A wet mount was prepared over a glass slide using the overnight culture Evaluation of cell chaining was performed using standard light microscopy at 1000x under oil immersion using an Axioscope 40 upright microscope (Zeiss).

To evaluate the intactness of the Tat quality control mechanism, DADE cells were freshly transformed with pPRATABC containing desired *tatB* point mutations and pSALect α3. Transformants were plated on LB agar containing aTc and Cm. Individual colonies were selected and grown overnight at 37°C in LB with aTc and Cm. Cultures were normalized to an OD_{600} of 1 using LB. Ten-fold serial dilutions of cells were created from 10^{-1} to 10^{-7} using LB and appropriate amounts of normalized cells. LB agar plates were made with either aTc and Cm (control plate) or 300 µg/mL Carb. Diluted cells were spotted onto the plates (5 µL volume) and plates were incubated overnight at 30°C.

CHAPTER 4

ENHANCED PRODUCTION OF A GH5 FROM *FUSARIUM GRAMINEARUM* WITH SHUFFLED DISULFIDE BONDS AND ITS UNEXPECTED INTERACTION WITH DEXTRAN-CONTAINING MEDIA

Abstract

Cellulases aim to become the most widely produced enzyme class due to their use in conversion of lignocellulosic biomass into liquid fuels. However, large-scale production of these enzymes in natural fungal systems is difficult due to expensive and time-consuming culture methods. Common heterologous protein-production hosts, such as *E. coli*, are an attractive alternative choice to generate designer cellulase cocktails, but have proven challenging to work with due to low titer cytoplasmic expression and incomplete post-translational modification. Cellulases are naturally secreted from cells and fold in the endoplasmic reticulum. In this work we show how targeting a fungal endoglucanase to the periplasmic space of *E. coli* via the general secretory pathway can dramatically increase its production. Further, we evaluated an expression platform capable of forming disulfide bonds in the cytoplasm of *E. coli*, resulting in an almost 1,000-fold increase in cellulase titer. Lastly, we describe the novel interaction of the recombinant cellulase with dextran-containing media of a size exclusion column and speculate how this interaction could enhance the kinetics of the cellulase in nature. *This chapter is work in progress towards two manuscripts. The first describes the production*

increase of a recombinant cellulase due to disulfide bonding; it is targeted towards a protein expression and purification journal. The second describes elucidation of endocelluase interaction with dextan-containing size exclusion column.

Introduction

Conversion of biomass into liquid fuels is a key piece of the renewable-energy forecast due to cellulose being the world's most abundant source of terrestrial carbon. First-generation technologies are focused on conversion of lignocellulosic material to a fuel such as ethanol in four steps: growth/harvest of the biomass source, pretreatment to begin harvesting sugars and create pores in the cellulose fibrils, enzymatic hydrolysis to release sugars, and fermentation of the sugars to a desired fuel [62]. Further development of biomass-conversion technologies has focused on cellulases enzymes due to their relatively high cost of production, incompletely understood catalytic mechanisms, and limited reusability [60, 75]. As such, several design parameters must be considered when creating optimal cellulase cocktails including activity, production amount, stability at process conditions, synergism with other cellulases and recovery [62]. *T. reesei* and evolved strains of this fungus, such as Rut-30, have served as the main production platform for enzyme cocktails due to their protein secretion titers reaching 100 g/L [66].

However, the heterogeneity of biomass sources has necessitated further refinement of cellulase cocktails for more efficient and complete depolymerization [60, 62]. Bioprospecting studies of diverse organisms from fungal plant pathogens [71] to

anaerobic gut fungi [72] have served as sources of novel cellulases or cellulosome scaffolds to enhance degradation. One notable example is a GH61 protein, discovered by Novozymes, that when combined with the natural *T. reesei*-produced cocktail increased yield on barley straw [67, 135, 136]. The GH5 (glycosyl hydrolase family 5), that was engineered for increased production in Chapter 2 of this work, was found in a bioprospecting study that showed enhanced degradation of a variety of biomass feedstocks such as switchgrass by fungal plant pathogen extracts over conventional cellulase cocktails [71]. Combining enzymes to form optimal cocktails for each biomass source and pretreatment strategy is key to increasing yield and lowering cost of lignocellulosic-based fuels [62]. Unfortunately, development of heterologous protein secretion in fungal hosts has been difficult due to low knowledge about the organism, limited genetic tools, and difficult culturing techniques [16, 72-74, 137]. Even *T. reesei*, with all its potential for high-secretion titers of its own enzymes, has had limited success secreting recombinant proteins [16, 66, 73, 138].

Platform expression hosts such as *E. coli*, *S. cerevisiae,* and *B. subtilis* are being developed as heterologous production organisms for these designer cellulases [74]; however, these hosts have several disadvantages compared to natural hosts. The vast majority of fungal or bacterial cellulases are secreted. As such, fungal cellulases begin their journey as an unfolded polypeptide chain being translocated into the endoplasmic reticulum by the Sec61 machinery [41]. Protein folding occurs in the endoplasmic reticulum amongst chaperones to ensure proper disulfide bonding, the beginning of N-linked glycan attachment, phosphorylation and complex formation [41]. The cellulase is

sent to the golgi apparatus where N-linked glycan refinement and O-linked glycosylation occurs, followed by secretion from the fungus [41]. These critical post-translational modifications can affect cellulase activity, stability, or specificity [139]. Performing all these necessary post-translational modifications is a current limitation in heterologous hosts and is a main contributor to the lack of success of producing highly active cellulases from these organisms [16, 73, 74].

For instance, disulfide formation typically occurs in the *E. coli* periplasm and is not favored in the cytoplasm due to the presence of thioredoxin and glutaredoxin/glutathione pathways that keep proteins in their reduced form [12]. To circumvent this limitation, *E. coli* strains have been developed with genetic mutations to alter the reducing potential in the cytoplasm to favor disulfide bond formation [140]. Additionally, strains were created that added the enzyme disulfide bond isomerase (DsbC) to the cytoplasm to accompany the potential to form disulfide bonds [140]. DsbC is a periplasmic-chaperone that functions to rearrange disulfide bonds to be formed between proper pairs of cysteines since these bonds are often formed between incorrect cysteines after translation [122]. Further refinement of cytoplasmic disulfide-bond producing strains led to the commercial strain SHuffle, which has been shown to be able to produce tissue plasminogen activator with its six disulfide bonds as well as a Cel9A with three disulfides at a titer of 250 mg/L [12]. We plan to use this commercial strain to overexpress the Cel5A from *F. graminaerum* as it potentially has three disulfides.

Results

Cel5A production is enhanced by disulfide formation. Cel5A contains six cysteine residues that are predicted to form three disulfide bonds based on the homology model of the CBM and catalytic domain. The expected disulfide pairings are: C8-C25, C19-C35, and C273-C310. Many cellulases, including those produced from fungal plant pathogens, are secreted from cells so that they may act on foreign plant material [141]. As such, they typically go through the general fungal secretion pathway where they fold in the endoplasmic reticulum enabling the formation of disulfide bonds [41]. We tested the production of Cel5A in SHuffle T7 express to determine the influence of disulfide bonding on expression in *E. coli* (**Figure 4.1a**). Since the SHuffle strain contains a cytoplasmic DsbC, it can rearrange the not sequential disulfide bonds in the Cel5A to form bonds between proper cysteines [12].

We found a dramatic increase in production of Cel5A produced in SHuffle T7 express compared to BL21(DE3) cells. Previously, we estimated the production of Cel5A wt to be about 0.02 mg/L. In SHuffle, the protein was overexpressed and at a high percentage of the total protein produced (**Figure 4.1b**), which is typically observed for other well expressed proteins in *E. coli*. Optimization of media (terrific broth), expression time (24 hours), and production temperature (25°C) provided even greater expression (data not shown). A rough estimate of the production of Cel5A in SHuffle T7 express was between 10 to 20 mg/L, almost a 1,000-fold increase over previous cytoplasmic expression. However, production of Cel5A in SHuffle was much lower than what was previously reported for a Cel9A [12], probably due to the need for further

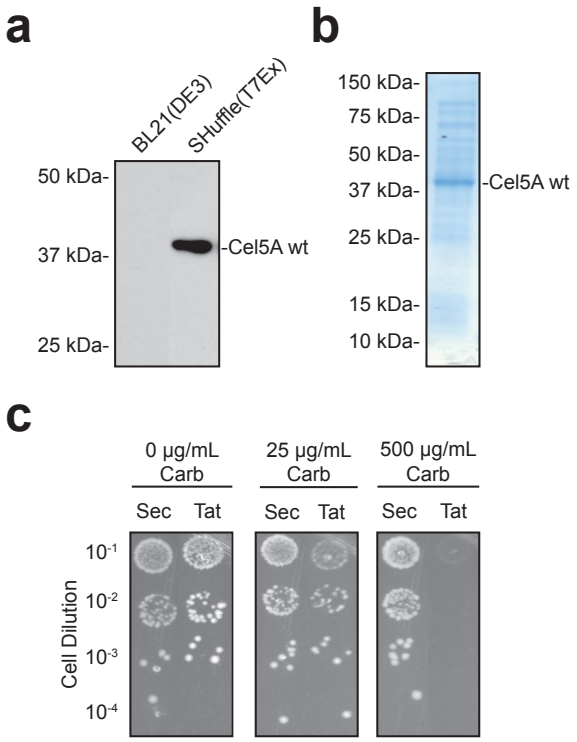

Figure 4.1: Cel5A wt production with disulfide bonds. A) Comparison of Cel5A wt production in BL21(DE3) and SHuffle T7 express (T7Ex) strains. SHuffle stains contain genetic mutations to allow for disulfide bonding to occur in the cytoplasm as well as produce DsbC. Lanes were loaded with 3 µg total protein and Cel5A bands were detected using an anti-Flag antibody. B) Soluble cell lysate for Cel5A wt produced in SHuffle T7 express and stained with Coomassie blue. C) Comparison of growth phenotypes when Cel5A wt is used in the Tat- or Sec-based folding reporters. The Tat-based reporter is pSALect as described in chapter 2. The Sec-based folding reporter is described in [49] and contains a N-terminal DsbA signal sequence and C-terminal Bla.

optimization of expression conditions, use of autoinduction media and fermentation in a bioreactor.

Cel5A was isolated from a study of fungal cell extracts on various biomass sources [71]. As such, it is likely to be a secreted protein that enters the endoplasmic reticulum via the Sec61 translocon. Additionally, evidence herein suggests disulfide-bonding plays a key role in the production of Cel5A. We decided to test the production

of the cellulase using a periplasmic folding reporter that utilizes the general secretory pathway (Sec) of *E. coli* in comparison to the previously used Tat-based reporter (**Figure 4.1c**). We found that when Cel5A is targeted to Sec, it can grow at a much higher concentration of carbenicillin (> 500 μg/mL Carb) than when translocated through Tat (25 μg/mL Carb). The growth resistance phenotype is likely due to an increase in production because of proper disulfide bond formation in the periplasm due to the oxidizing environment and chaperones to promote folding.

Cel5A catalytic domain is less soluble than the CBM. Although Cel5A is a single polypeptide chain, like many cellulases it contains two distinct domains connected by a flexible linker [60]. Since the two domains have discrete functions and are separated, we tested their relative solubility individually in the Tat-based folding reporter (**Figure 4.2**). When the CBM was produced by itself, it gave rise to a higher resistance growth phenotype than the catalytic domain. This is not surprising since the mutations discovered in the earlier genetic selection and screen were both found in the catalytic domain. The only mutation obtained in the CBM caused a moderate increase in production. Further, full-length Cel5A showed highest sensitivity to carbenicillin. These results provide a recommendation for a more focused random mutagenesis strategy in the future to target the engineering of less soluble domains and potentially to not alter function of regions that are already produced well in *E. coli*.

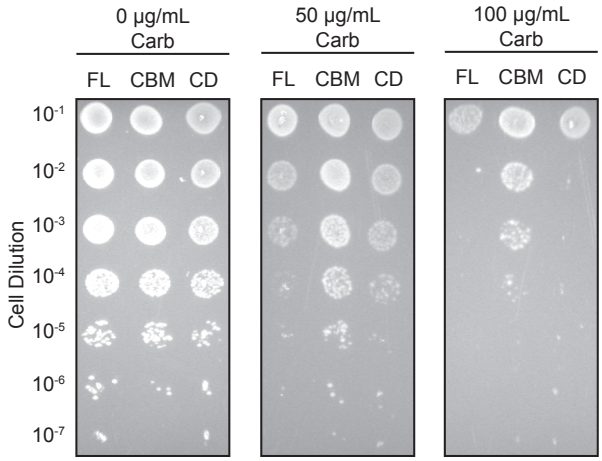

Figure 4.2: Deconstruction of Cel5A domains. Comparison of the growth phenotypes for full length (FL), the carbohydrate binding module with linker (CBM) and catalytic domain (CD) of Cel5A wt using the Tat-based folding reporter.

Discovery of cellulase interaction with dextran-containing media. Previously we have shown by size exclusion chromatography that Cel5A wt and Cel5A TD elute as single peaks and suggesting that a loss of aggregation could not explain the production difference observed for the proteins (**Figure 2.7d**). However, when the molecular masses of the proteins are compared to other globular protein standards using the same Superdex 75 column (**Figure 4.3a**), it was found that the proteins eluted at an expected molecular mass between 12-16 kDa. The molecular mass for Cel5A is predicted to be approximately 43 kDa and it has been observed at that molecular mass after separation by denaturing and reducing SDS-PAGE. The Cel5A wt protein was separated using a second size exclusion column containing Superdex 200 media and it was observed that the protein eluted at an even smaller molecular mass of

approximately 3-5 kDa (data not shown), which is outside of the linear separation region for the column.

We hypothesized two reasons why a protein may elute at a smaller volume than expected from size exclusion chromatography. First, the shape of the protein maybe conducive to diffusion into pores. Proteins used as standards are assumed to be spherical shaped; however, rod-shaped proteins might be able to penetrate deeper into the size exclusion media. Second, the protein could be interacting with the Superdex media, which is composed of crosslinked dextran and agarose. Binding to Superdex would cause the proteins to elute at a smaller than expected molecular mass. Dextrose, like cellulose, is a polymer of glucose. Dextrose contains α(1,6) and α(1,3) linkages, with the α(1,3) determining the degree of crosslinking. In contrast, cellulose contains a β(1,4) linkage. Searching through the carbohydrate-active enzymes (CAzy) database did not yield any CBMs that are known to interact with dextrose or agarose [99].

Conveniently, our genetic selection and screen isolated C5A.1 that contained a mutation in the CBM and loss of activity on insoluble substrates (data not shown). Interaction with CMC is not dependent on a functional CBM, while binding to insoluble substrates such as Avicel or BMCC is [60]. Additionally, though conservation analysis and homology models, we predicted that the CBM mutation of G10D could have a negative impact on the CBM structure, resulting in a lower activity on insoluble cellulose substrates. If the G10D mutation has a negative impact on CBM binding, it could change how the protein interacts with the Superdex media (**Figure 4.3b**). We found that both the C5A.1 and C5A.2, which contain the G10D mutation, eluted with peaks at

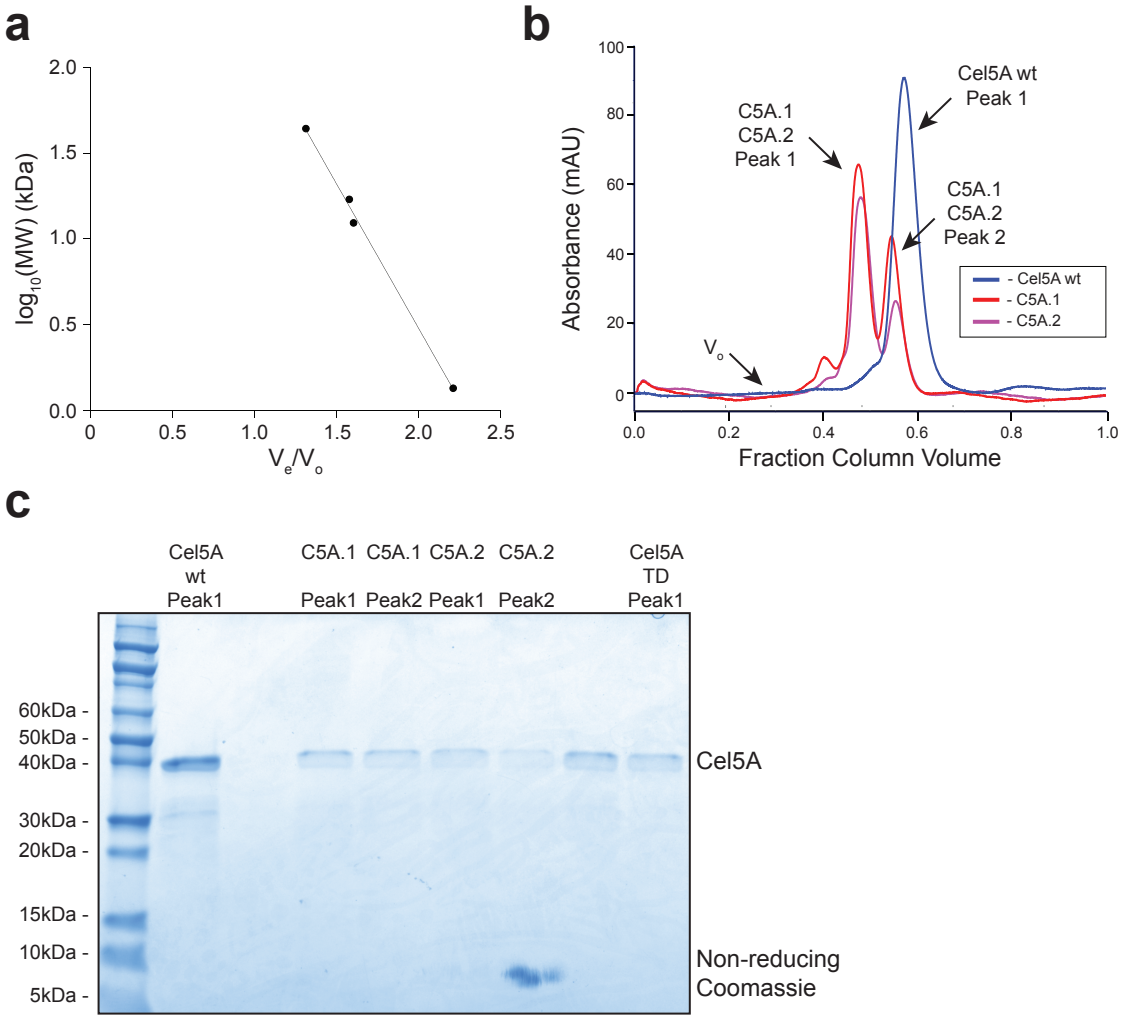

Figure 4.3: Size exclusion chromatography of C5A.1 and C5A.2. A) Protein standards separated using a Superdex 75 size exclusion column. The void volume (V_o) was 8.25 mL and V_e refers the elution volume (center of the peak) for each standard. One column volume is 24 mL. The y-axis is the $\log_{10}$ of the molecular weight (MW) of the monomeric standards. A standard curve was generated by fitting a straight line to the $\log_{10}$(MW) as a function of V_e/V_o for the protein standard set within the linear range for Superdex 75 media. B) Absorbance profiles for Cel5A wt, C5A.1, and C5A.2 separated using a Superdex 75 size exclusion column. **Figure 2.7d** shows the absorbance profile for Cel5A TD. C) Proteins from size exclusion chromatography peaks were denatured (not reduced) and separated by SDS-PAGE followed by staining with Coomassie blue. For C5A.1 and C5A.2, peak 1 refers to the peak at ~4.75 cv and peak 2 at ~5.5 cv.

higher molecular masses than the Cel5A wt or Cel5A TD. The estimated molecular mass of the left-shifted peak is approximately 34 to 36 kDa, which is closer to the expected molecular weight of Cel5A. Protein samples from each of the peaks were denatured, separated by SDS-PAGE and stained with Coomassie blue (**Figure 4.3c**). Even though the proteins separated differently by size exclusion chromatography, they all migrated at the same molecular mass when denatured. The proteins were not reduced to remove the possibility of intermolecular disulfide bonds causing shifts in apparent molecular mass. Taken together this suggests that Cel5A is interacting with the Superdex column through its CBM, which can be mutated to lower its binding.

BioSAXS predicts overall Cel5A structure. Based on size exclusion chromatograph of Cel5A with and without the G10D CBM mutation, it seems likely that the protein is interacting with the Superdex media causing a shift to an apparent lower molecular mass. However, protein shape can affect its diffusion through pores. Light scattering can be used to determine overall protein shape, providing a radius of gyration and estimated molecular mass, to ensure the proteins eluting at different volumes by size exclusion chromatography do in fact look similar. Using a higher energy wave source such as x-rays, it is possible to predict the envelope of the protein [142] to gain information on the orientation of the CBM and catalytic domains, since they are unable to be co-crystalized due to their flexible linker. As such, biological small-angle x-ray scattering (BioSAXS) has been used to predict the structures of three such cellulases [143-145]. BioSAXS has the advantage over crystallography in that it can be performed

in solution and captures the multiple states a protein may be in; however, it is low resolution and can only be used to predict an ensemble envelope. It was observed that the proteins had a “tadpole”-like structure with the CBM located far from the catalytic domain with the connecting flexible linker to allow the endocellulases to beak multiple bonds through one binding event [143, 144].

BioSAXS measurements are highly sensitive to protein concentration and aggregation, especially for the smallest angle measurements. Measurements were made over a wide concentration range to ensure the protein was not aggregating. Cel5A TD was used to perform the concentration series measurements since it was obtained at the highest concentration. It was observed that the protein aggregated in solution at 42.5 μg/μL by the sharp increase in intensity for small angle scattering (**Figure 4.4a**) and was found to aggregate until a concentration of about 1.28 μg/μL was reached (**Figure 4.4b**). It was also determined that the limit of detection was approximately 0.25 μg/μL where the signal to noise ratio was low (data not shown). Using these conditions, it was possible to compare the radius of gyration for C5A.2 after separation by size exclusion chromatography. C5A.2 peak 1 was estimated to have a radius of gyration of 26.092 ± 4.995 Å. Unfortunately, peak 1 for Cel5A TD did not yield usable SAXS data due to aggregation. Cel5A TD and C5A.2 samples not purified by size exclusion chromatography were found to have a radii of gyration of 30.625 ± 3.578 Å (**Figure 4.4b**) and 31.453 ± 5.290 Å (**Figure 4.5a**), respectively. This very preliminary evidence suggests that the proteins do in fact have a similar radius of gyration and

spherical shape; however, these experiments would need to be repeated to ensure their validity and to decrease the error associated with the estimated radius of gyration.

In addition to size prediction, BioSAXS is a powerful method to predict the overall envelope structure of the protein. The distance distribution was obtained for Cel5A TD (**Figure 4.4c**) and a Kratky plot was created to show the protein was well folded (**Figure 4.4d**) [146]. Even though the data for the Cel5A TD had a better signal to noise ratio at high concentrations (**Figure 4.4e**), its small angle peak aligned well at both high and low concentrations suggesting the low concentration data could be used to predict an envelope. Further, the wide-angle data sloped upward suggesting the protein contains an unstructured region, probably the linker portion. The DAMMIF algorithm was used to predict ten envelope structures and fit the data with a χ^2 of 0.110 for the DAMMIF fit shown (**Figure 4.4b**) [147]. Each envelope showed a "tadpole" structure with a large catalytic domain disc connected to a thin linker and larger CBM. Envelopes that showed similarity (9 out of 10) were combined to form one final structure using DAMAVER. Due to the linker being able to adopt numerous structures, that portion of the protein was predicted to be larger than observed for any singular prediction after averaging the structures [148]. To remedy this problem, the DAMFILT algorithim was used to remove loosely connected atoms as well as those that were unlikely to occur in the linker region to form the final envelope structure (**Figure 4.4f**). The catalytic domain and CBM were superimposed on the envelope with a hypothesized linker connecting the two. As was observed previously, the protein does adopt a "tadpole" structure with the CBM located far from the catalytic domain; although, I would rather refer to it as a

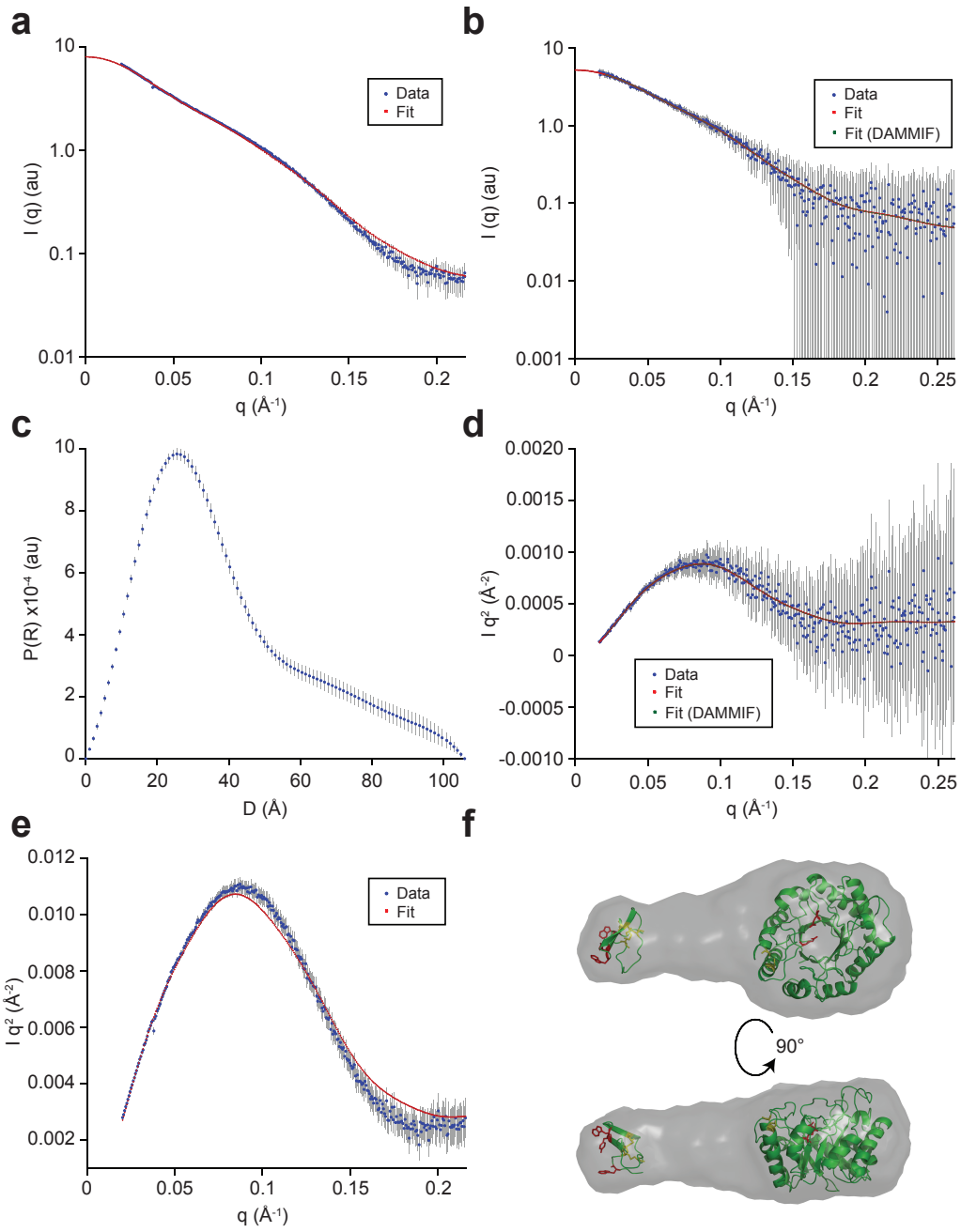

Figure 4.4: BioSAXS prediction of Cel5A TD envelope. A) Scattering intensity (I) as a function of scattering angle (q) for a 42.5 µg/µL sample of Cel5A TD. The scattering intensity is the average of 10 spectra subtracted from 10 blank spectra. Intensity is shown in blue and error as grey bars. The best-fit curve is shown in red. B) I vs q for a 1.28 µg/µL sample of Cel5A TD. In addition to the best-fit for the scattering data, the DAMMIF fit is shown in green. C) Probability distance distribution (P) as a function of particle diameter (D) predicted by primus (ATSAS) for a 1.28 µg/µL sample of Cel5A TD. Dmax was set to 106 Å. D) Kratky plot (I*q2 vs q) for a 1.28 µg/µL sample of Cel5A TD to show the quality of the data and fit. E) Kratky plot for a 42.5 µg/µL sample of Cel5A TD. F) Averaged DAMMIF predicted envelope (using 1.28 µg/µL sample) of Cel5A TD (grey) with superimposed CBM and catalytic domain homology models (green). Catalytic residues and those involved in cellulose binding are shown in red and predicted disulfide bonds in yellow.

"drumstick"-like structure. Due to the large separation of the CBM from the catalytic domain, it suggests that these portions maybe separated for future engineering efforts as discussed above.

A similar strategy was used to predict the envelopes for C5A.2 for a sample that was not subjected to size exclusion chromatography (**Figure 4.5**). The DAMMIF fit shown in **Figure 4.5a** fit the observed data with a χ^2 of 0.135. When the distance distributions were calculated, the protein was estimated to have a smaller D_{max} than was observed for Cel5A TD and the two distributions looked qualitatively different (**Figure 4.5b**). Additionally, an upshift in the wide-angle scattering data was not observed on the Kratky plot suggesting the loss of the unstructured linker region (**Figure 4.5c**). Envelope prediction was not as consistent as was observed for Cel5A TD and three proposed envelopes that were obtained numerous times are shown (**Figure 4.5d**). The envelopes predicted for C5A.2 qualitatively look very different than those for Cel5A TD, there is not a spread out linker region and the structure is more compact. Some of the models possibly contain two catalytic domains and could be a dimer. Homodimerization of C5A.2 would explain the mass shift observed in the size exclusion chromatogram. It seems that the C5A.2 has folded upon itself, possibly due to incorrect disulfide bond pairing. Additionally, envelopes were predicted for C5A.2 peak 1 (purified by size exclusion chromatography) with similar prediction of a more globular protein that is potentially a dimer (data not shown). Unfortunately, Cel5A wt and C5A.1 envelopes could not be predicted due to aggregation. The change in shape between C5A.2 and

Cel5A TD could explain the size exclusion results without CBM dextran binding leaving open both hypotheses to explain these unexpected findings.

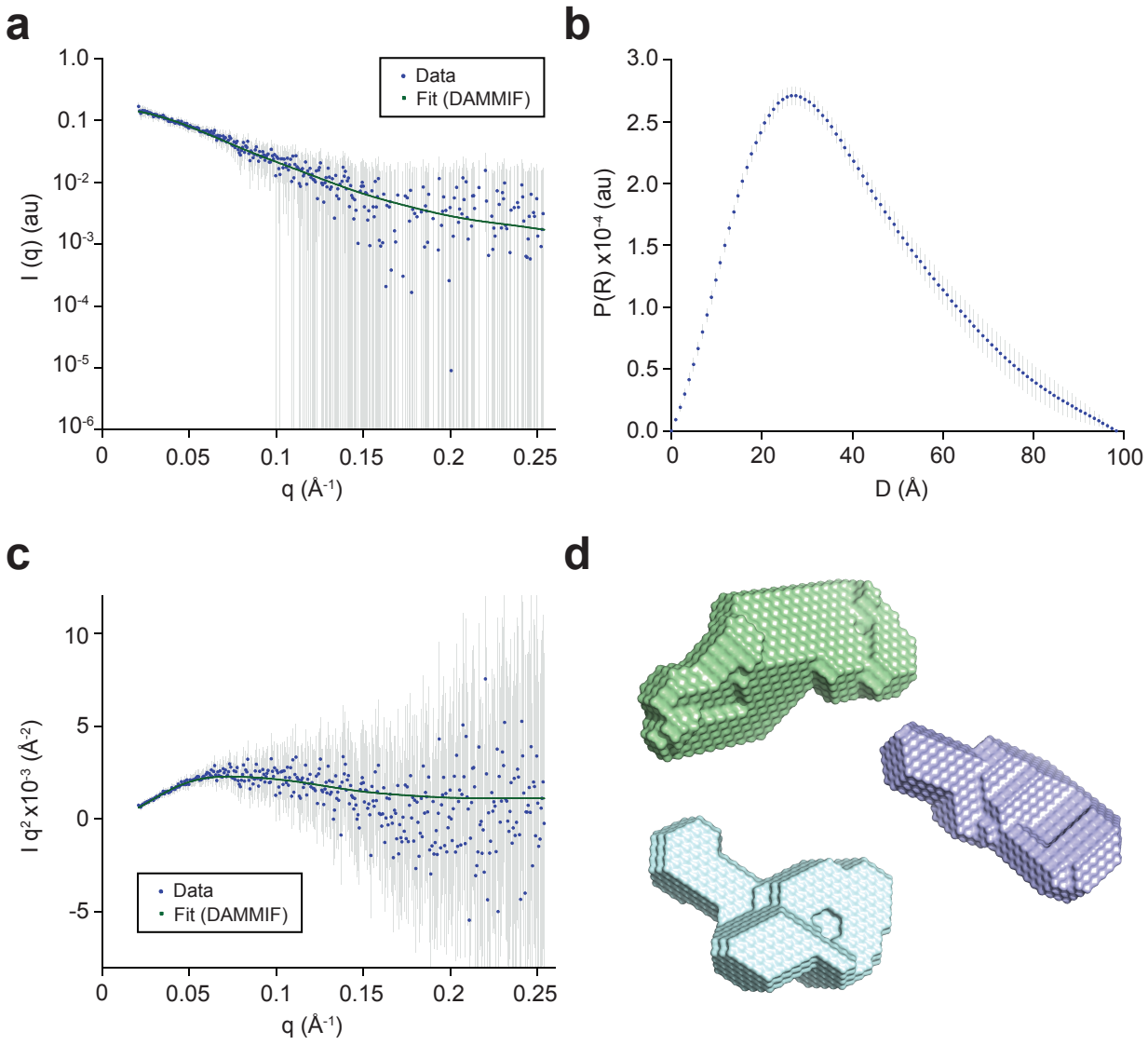

Figure 4.5: BioSAXS prediction of C5A.2 envelopes. A) Scattering intensity as a function of scattering angle for a 0.98 µg/µL sample of C5A.2 and a representative DAMMIF fit. B) Probability distance distribution as a function of particle diameter for C5A.2. C) Kratky plot for C5A.2. D) Three representative DAMMIF predicted envelopes for C5A.2. The structure shown in green was the most often predicted.

Discussion

Heterologous cellulase production holds promise in creating designer cocktails or cellulosomes to degrade the wide range of biomass sources to be grown as a renewable energy source [60, 62]. While fungi are the ideal production host, lack of genetic characterization or culturing methods has necessitated moving expression into model organisms [73, 74]. However, this has been met with challenges due to the differences in production and secretion pathways between organisms [16, 74]. We have overcome these issues by translocating Cel5A through the Sec pathway and by producing it in a strain engineered for disulfide bond formation. The ability to form disulfide bonds in Cel5A raised its expression a predicted 1,000-fold, even though these disulfide bonds are unnecessary for activity. Correct disulfide rearrangement is predicted to be critical to this enhancement in production due to the presence of DsbC in the cytoplasm. Additionally, targeting the Cel5A to the Sec pathway, which is homologous to the protein secretion pathway in fungi, provided more translocated product than through the Tat pathway

Overcoming disulfide bonding limitations is only problem with heterologous production of cellulases. Another important post-translational modification to cellulases is the incorporation of N- and O-linked glycans [139]. Currently, glycosylation is limited or results in the incorrect incorporation of glycans for all heterologous cellulase hosts [16, 73]. Even one of the best evolved hypersecreters of cellulases, Rut-30, contained a mutation in the glycosylation machinery resulting in the improper glycoform addition [66]. Further, overexpression of cellulases and media formulation can lead to incorrect

glycan attachment [73, 149]. Glycans are thought to play multiple roles for cellulases including stabilizing affects and promoting interaction with carbohydrates [139]. O-linked glycans in the CBM and N-linked glycans in the catalytic domain are hypothesized to enhance binding to cellulose and may play a role in substrate recognition. As such, correct glycoforms must be added to ensure only desired interactions occur [139]. Glycosylation can provide thermostability for the protein, decrease aggregation and O-linked glycans in the linker region are thought to eliminate protease cleavage sites [73, 139]. Additionally, glycans can provide rigidity to linker regions to change the overall shape of the cellulase, alter its processivity, and possibly even provide allosteric control [73, 139]. As we found through our BioSAXS data, the Cel5A linker does separate the CBM from the catalytic domain; however, if it was linearly stretched by incorporation of O-linked glycans, the separation would be even longer. Lastly, it is believed that extracellular glycan trimming plays a crucial role in obtaining the desired glycan in each position [149]. This interest in glycosylation of cellulases for enhance stability, activity, and synergism has effected the evaluation of heterologous hosts for production as well as how to engineer these enzymes. Organisms such as *E. coli* have the advantage of not containing N-linked glycosylation machinery possibly enabling building of specific glycans to add to cellulases without needing further processing.

Many of the lessons learned for the heterologous expression of cellulases must also be applied to their engineering. Apart from fungal-strain engineering for enhanced cellulase production [66, 73], the majority of the protein engineering approaches for

cellulases have come from model organisms [77]. In this chapter, we have shown how formation of disulfide bonds and translocation through the Sec pathway can enhance production, which are both hypothesized to naturally occur for Cel5A. Additionally, we discovered that the catalytic domain is produced at a lower level than the CBM in *E. coli*, suggesting future engineering attempts to enhance production should be focused on that domain. Due to an incomplete mechanistic understanding of cellulases and a lack of detailed assays to describe cellulases, it is challenging to engineer them without compromising function, stability, or synergism [60]. Some of the overlooked cellulase characteristics that can contribute to their function include: shape, glycosylation pattern, protein-protein interfaces and allosteric regulation. Future engineering attempts must be cognizant of all of the factors that can affect cellulase performance to create careful screens or selections for enhancement of specific attributes while focusing on the main goal of biomass depolymerization.

Lastly, through characterization of Cel5A, we discovered an enigmatic separation of the protein by size exclusion chromatography where the protein elutes at a smaller apparent size than was observed through separation by SDS-PAGE. Through mutation of the CBM, the protein was found to partition to a higher molecular mass which prompted the hypothesis that the CBM is interacting with the Superdex media, probably due to dextran binding [150]. Although, CBM binding has not been listed in CAzy, several enzymes do show dextranase activity [99]; however, none are known from the GH5 family. From a related *Fusarium* species, a secreted endodextranse has been characterized, which suggests possible conservation among the similar organisms

[151]. An equally intriguing hypothesis for the observed size exclusion profile is the shape of the protein influencing diffusion through Superdex pores. It was found that the Cel5A TD has a "tadpole" shape as found for other endocellulases [143, 144]. However, C5A.2 does not show an extension of the CBM from the catalytic domain and has a more globular shape than the Cel5A TD. Although the BioSAXS data is preliminary and further controls such as catalytic domain only, Cel5A G10D, or Cel5A with CBM mutations to hydrophobic residues need to be tested by size exclusion chromatography, these results may provide answers on how shape may affect cellulase catalysis. Although only speculation at this time, the rod-shaped cellulase may diffuse better into pores than a globular protein, which becomes another cellulase design constraint and critical when analyzing their potential for biofuel production.

Methods

Strains and plasmids. BL21(DE3) (Novagen) and SHuffle T7 express cells (NEB) were used for production of Cel5A from a pET28a plasmid containing an N-terminal His-Flag tag as described in Chapter 2. MC4100 cells were used in spot plating experiments. pSALect is the plasmid containing the Tat-based folding reporter, and was previously described in [58, 125] and Chapter 2. pDMB is a Sec-based periplasmic folding reporter that consists of the signal sequence from *dsbA*, a multiple cloning site, and *bla* as described in [49]. pDMB Cel5A wt was created excising *cel5A* from pSALect Cel5A wt with XbaI and SalI, and adding it to pDMB with the same cutsites. The Cel5A CBM with linker ($cel5A_{CBM\text{-}L}$) was PCR amplified using the following two oligoneucleotide

primers: 5'- cgacgtgcgactgcgtctagacaatccagcgcctgggctc and 5'-gaggaggtcgacatcacccggggtgccg. The catalytic domain of Cel5A (*cel5A*$_{CD}$) was PCR amplified using the following two oligoneucleotide primers: 5'-gaggagtctagaggcaagttcttgtgggcgggtg and 5'-ccagcgtttctgggtggtcgacgacataggttttcagcaggctattgtag. The PCR-amplified *cel5A*$_{CBM-L}$ or *cel5A*$_{CD}$ were added to pSALect using the NdeI and SalI cut sites.

Protein expression and spot plating. Cel5A wt was produced from pET28a as described in Chapter 2. Spot plating experiments using pSALect or pDMB were carried out as described in Chapter 2.

Size exclusion chromatography. Size exclusion chromatography was performed using a Superdex 75 10/300 GL column (GE Lifesciences) and AKTApurifier FPLC (GE Lifesciences). The mobile phase was 50 mM sodium acetate (pH 5.0). Protein standards were from the gel filtration standard set (BioRad) and CytochromeC (Sigma). The standards within the linear range for Superdex 75 are: Ovalbumin (44 kDa), Myoglobin (17 kDa), CytochromeC (12.4 kDa) and Vitamin B_{12} (1.35 kDa). Thyroglobulin (670 kDa) and blue dextran (~2 MDa) were used to determine the void volume of the column. Cel5A proteins were previously expressed, purified and concentrated as mentioned in Chapter 2. Size exclusion chromatography was carried out with 500 µg Cel5A protein at 0.2 mL/min.

BioSAXS data collection and envelope construction. Proteins were purified as previously described. Extinction coefficients for each Cel5A protein were estimated using concentrations measured by BCA assay with a BSA standard and measured absorbance values. A dilution series of Cel5A TD was created between 40 µg/µL to 0.1 µg/µL since this was the most concentrated protein. Other proteins were tested at the highest concentration available after purification. The Cornell High Energy Synchrotron Source (CHESS) was used as the source of high energy X-rays and hutch G1 was set up with a robotic sampling unit as described in [152]. For each sample, 10 spectra were taken by exposing for 2 seconds and these spectra were averaged. Blank samples of 50 mM sodium acetate buffer were subtracted from each sample and scattering was collected while oscillating samples to reduce X-ray damage. After each sample was measured, the absorbance at 280 nm was collected to give an accurate concentration of the sample used to collect the SAXS data. Using the measured concentration and a Guinier fit ($I(q)$ vs q^2), it was possible to determine aggregation and estimate the radius of gyration for each sample.

The distance distribution was calculated using the Primus program from the ATSAS package (EMBL Hamburg) [146]. The D_{max} was set to 106 Å. The DAMMIF algorithm was used to predict ten envelope structures for Cel5A TD [147]. The predicted envelopes were reoriented to overlap using the DAMSUP algorithm and were compared using the DAMSEL program [148]. DAMSEL removed any structures that did not show similarity to each of the other structures generated. Remaining structures were combined using DAMAVER, and low occupancy or loosely connected atoms were

removed from the linker region by DAMFILT [148]. Likely position of the homology-modeled catalytic domain and CBM were predicted individually using the SUPCOMB program.

CHAPTER 5

SPECIFIC AND COVALENT CROSSLINKING OF PROTEINS IN *E. COLI* USING SORTASE A OCCURS WITH LIMITED EFFICIENCY

Abstract

Genetic fusion proteins are a major contributor to recombinant biotechnology today serving as reporters for folding, cellular trafficking markers, and bivalent therapeutics. Although genetic fusions are typically soluble and active, some fusion products do not fold properly due to interactions between the folding nascent domains. As such, we sought to create a robust method that would alleviate the problems associated with improper folding interactions of genetic fusions by covalently linking proteins after they are fully folded and active using the SortaseA (SrtA) enzyme from *Staphylococcus aureus*. We have used this technique of protein ligation to successfully crosslink a variety of proteins inside *E. coli* cells including fluorescent proteins (CyPet and YPet), enzymes (β-lactamase and enolase) and binding proteins (maltose binding protein). We also created a unique method of measuring the fusion of two proteins using Förester resonance energy transfer (FRET). This technology has implications for the creation of fusion proteins such as protease-resistant bifunctional antibodies, the construction of channeled multi-enzyme pathways, and the directed polymerization of proteins. *This chapter is a manuscript in preparation for submission to the Journal of Biotechnology.*

Introduction

The genetic fusion of proteins is the most common way to link the functions of two proteins together and is a widely used tool in biotechnology today [5]. However, such fusions are only possible as C- to N- terminal fusions and can suffer from co-translational misfolding problems [9, 10]. For instance, tandem scFvs, which are increasingly popular therapeutics due to their dual functionality [153], are prone to misfolding as genetic fusions and are largely produced in inclusion bodies in *E. coli* [154, 155]. Refolding proteins from inclusion bodies is laborious and does not guarantee the final product is folded accurately nor retains high activity. Numerous chemical reagents have been developed to allow post-translational assembly of proteins, but they often result in non-specific linkage, require the incorporation of novel functional groups or are transient. One method that has proved successful over the past few years is the enzymatic linkage of proteins [156, 157]. In particular, the enzyme SortaseA (SrtA) from *Staphylococcus aureus* [158] has stood out as the most popular choice for protein-protein ligation and has been extensively studied, see reviews [159-162]. Using SrtA and related variants, it has been possible to immobilize proteins on solid surfaces such as PEG [163] and Biacore chips [164] as well as label cells *ex vivo* with modified fluorophores [165, 166]. SrtA was used to create circular proteins by crosslinking their N- and C-termini, which increased the half-life of these proteins [167-169]. A related and calcium-independent SrtA has been used in yeast to create circular proteins and perform protein-protein ligation [170, 171]. Further, SrtA has engineered to expand its substrate specificity creating novel fusion targets [172, 173].

To date, enzymatic protein-protein ligation by SrtA has been primarily limited to *in vitro* applications [174] or use in eukaryotic cells [170]. However, to expand this technology to post-translational protein fusion, it is desired to fuse proteins inside *E. coli* since it is the preferred host for recombinant protein production. Problems associated with SrtA-based protein fusion inside *E. coli* have been attributed to low calcium ion concentration [159, 170] and the high K_m value of SrtA for its targets [175-177]. It is well established that the presence of calcium enhances the kinetic turnover for SrtA by upwards of four-fold [178]. Recently, an engineering study used rational design to remove this dependence, creating a SrtA variant whose binding kinetics was not affected by the presence of calcium [179]. The crosslinking reaction by *S. aureus* SrtA occurs in two steps [175], first it recognizes a peptide recognition sequence (LPXTG) found at the C-terminus of the desired protein to be fused and hydrolyzes the peptide bond between the threonine and glycine residues [180]. The second step is the ligation of the threonine residue to an amine group as part of a pentaglycine motif found on a Lipid II carrier via a transpeptidation reaction. Many studies have taken advantage of the fact that the motif on the peptidoglycan carrier is identical to an N-terminal glycine amino acid residue to create methods of protein-protein fusion [159]. Since the SrtA reaction occurs in two steps, the enzyme has a K_m for both the recognition sequence and glycine repeats, which have been measured *in vitro* to be 141 μM and 24 μM, respectively [175]. However, other methods have reported K_m values of upwards of 8.76 mM for LPXTG peptides and 140 μM for the GGG substrate [181]. Since few proteins accumulate to the high micromolar or low millimolar level inside cells, fusion of

proteins may be kinetically limited and occur to a low degree. Other engineering efforts have created a SrtA variant with a 30-fold higher K_m for its recognition sequence; however, at an almost equal fold decrease in K_m for a triple glycine acceptor [181].

Even though problems of *in vivo* protein fusion using SrtA in *E.* coli have been hypothesized, we sought to create an optimized strategy for post-translational protein-protein fusion. To that end, we have developed a simple crosslinking strategy based on intein cleavage to unmask an N-terminal triple glycine motif and render the protein competent for fusion via SrtA (**Figure 5.1a**). Inteins, or intervening proteins, are a class of proteins that can self-excise from a peptide chain [182]. Inteins have been engineered to cleave themselves from the N-terminus of a protein to create a label free method of protein purification [183]. We use the same strategy here with a constitutively active intein, to create a protein with three N-terminal glycines. Additionally, inteins are highly soluble and may enhance the production of the protein to be fused. The intein-containing protein is co-expressed with its desired fusion partner containing a C-terminal LPETG SrtA recognition sequence and the SrtA enzyme. Protein-protein ligation is found to be robust for highly expressed recombinant proteins due to the production amount overcoming the kinetic limitations of SrtA. Furthermore, crosslinking of fluorescent proteins enables the use of FRET to measure the crosslinking reaction in real-time, potentially providing answers to the open question of SrtA binding kinetics for protein substrates.

Results

Fluorescent protein ligation inside *E. coli*. To test the hypothesis that SrtA can crosslink recombinant proteins *in vivo*, the SrtA recognition tags were genetically added to yellow and cyan fluorescent proteins YPet and CyPet [184], respectively. A ten amino acid flexible linker $(GGGSG)_2$ followed by the LPETG SrtA recognition sequence and a six-histidine tag were added to the C-terminus of YPet. The linker is present to prevent the recognition sequence from interfering with the folding of YPet and to reduce any steric hindrance that might interfere with SrtA binding. In addition to the LPETG tag, a non-specific control was created by replacing the recognition sequence with GGGSG residues. For CyPet, a Flag-epitope tag, ΔI-CM mini-intein [182, 183, 185], GGG tag and short linker (LSGG) were added to its N-terminus. The intein cleavage site is directly before the GGG tag. In nature, a pentaglycine motif is bound by SrtA, but it has been found *in vitro* that SrtA can crosslink proteins with single glycine [167]; however, most studies use GGG crosslinking tags since SrtA is thought to recognize this sequence better than a single glycine and further improvement was not obtained by adding more glycines [162, 175]. A control CyPet fusion was created by replacing the GGG tag with AAA residues that is known not to bind or be crosslinked by SrtA [162]. Fluorescent proteins were polycistronically produced from a high-copy, arabinose-inducible plasmid. A second IPTG-inducible plasmid containing *srtA* was created to produce the crosslinking enzyme.

Upon co-expression of the two fluorescent proteins with appropriate tags and the SrtA enzyme, YPet-CyPet ligation could be observed (**Figure 5.1b**). Importantly, removal

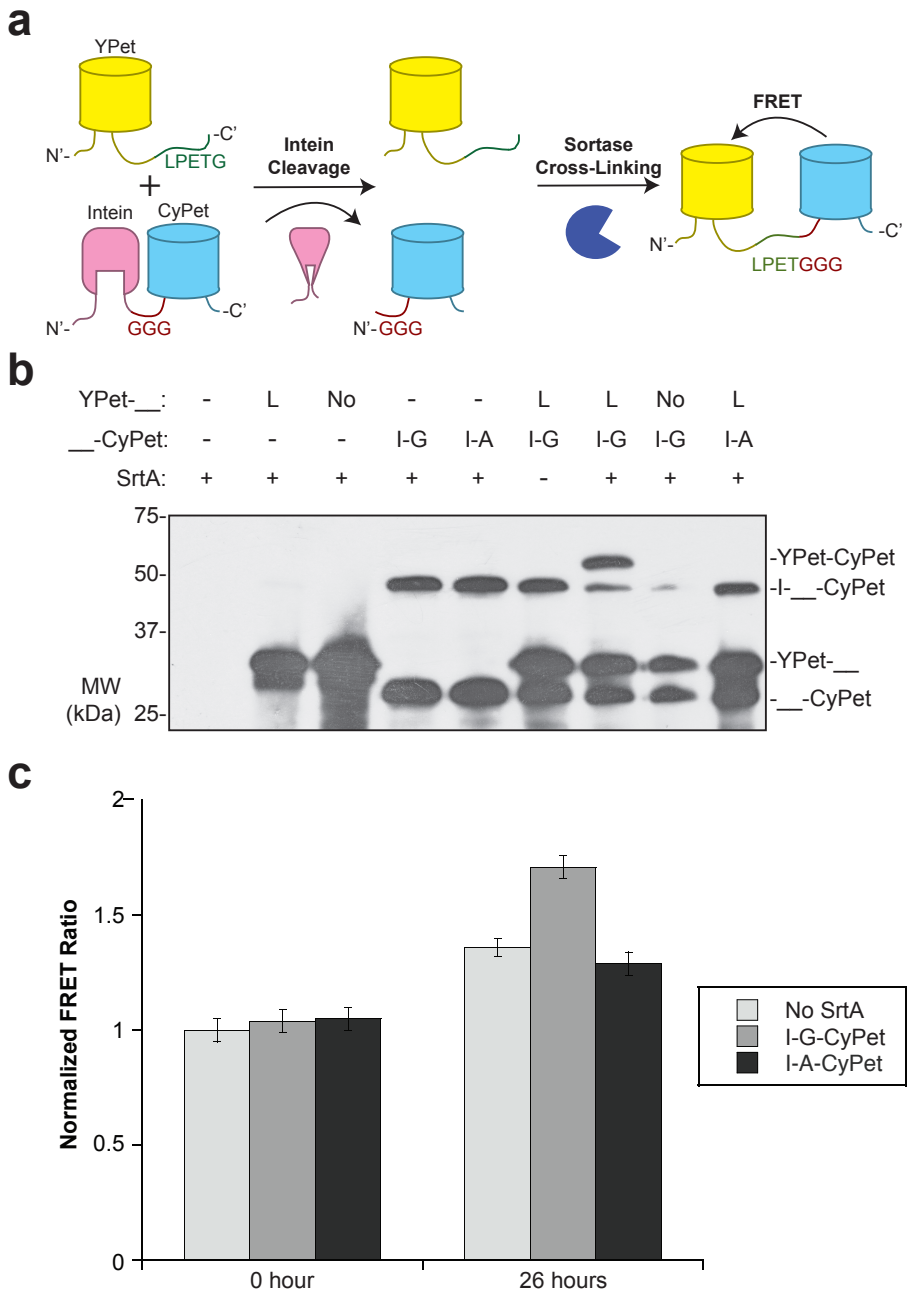

Figure 5.1: Post-translational ligation of fluorescent proteins in *E. coli*. A) Schematic of the SrtA fusion technique developed in this study. The intein cleavage occurs without induction and leads to an N-terminal GGG sequence. The SrtA enzyme recognizes the LPETG motif on the second protein, cleaves the peptide bond between the threonine and glycine residues, and covalently attaches a protein containing an N-terminal glycine following the threonine. Upon crosslinking of the fluorescent proteins CyPet and YPet, FRET can be used to measure the fusion event. B) Western blot of the in vivo crosslinking reaction of YPet and CyPet using SrtA. The cells in the SrtA (+) lanes have a SrtA-containing plasmid, while (-) have an empty plasmid. For YPet-__, (-) refers plasmid without the yPet gene, (L) is the SrtA recognition tag LPETG, and (No) is a GGGSG tag. For __-CyPet, (-) refers to plasmid without the cyPet gene, (I-G) is the intein followed by GGG residues, and (I-A) is the intein followed by AAA residues. The western blot was probed with anti-GFP that recognizes both YPet and CyPet. C) FRET ratio observed in E. coli before (0 hour) and after the crosslinking reaction (26 hour). No SrtA (light grey bars) refers to negative control cells without SrtA present that contain pOKD4 (empty); pET21a YPet-LPETG, Intein-GGG-CyPet. I-G-CyPet (medium grey bars) refers to cells that contained pOKD4 SrtA; pET21a YPet-LPETG, Intein-GGG-CyPet. I-A-CyPet (dark grey bars) refers to switching the GGG recognition sequence for AAA, and cells with pOKD4 SrtA; pET21a YPet-LPETG, Intein-AAA-CyPet. FRET ratio is the FRET signal (Ex: 430 nm, Em: 527) divided by the CyPet signal (Ex: 430 nm, Em: 527). Data is normalized to the No SrtA sample at 0 hours, and the error bars represent the propagated error from the standard deviation of 3 biologically separate samples.

of SrtA, mutation of the SrtA-recognition sequence, or changing the GGG binding residue to AAA eliminates the ligation product. Additionally, non-specific crosslinked products were not observed when the proteins were produced singularly with SrtA suggesting native proteins with appropriate SrtA recognition sites are produced at too low a level to be crosslinked. The tags on the CyPet protein were mutated to contain only one glycine residue to determine if single glycine was sufficient for crosslinking (**Figure 5.2**). It was found that YPet to CyPet ligation occurred similarly for the GGG and GAA tags, providing evidence that the SrtA only needs a single N-terminal glycine for crosslinking. Additionally, a control tag of AGG did not show any crosslinking (**Figure 5.2**), proving that intein cleavage is specific to the desired location directly before the fusion sequence.

These results show the specific and covalent fusion of two proteins in *E. coli* using SrtA; however, the amount of crosslinked product represents a small fraction of the entire pool of precursor proteins, about 12%. The amount of fusion observed (**Figure 5.1b**) was higher than many previous attempts due to proper plasmid combinations and co-optimizing the expression and crosslinking conditions. The fluorescent proteins were expressed at low temperatures (22°C) to maximize their production. Following protein production, high temperature incubation (37°C) was used to promote intein cleavage and SrtA catalyzed ligation. Even with the high temperature incubation step, it is observed that the intein cleavage is not complete, which lowers the pool of available GGG-CyPet. The rationale for choosing the YPet and CyPet fluorescent proteins was to measure the amount of crosslinking by FRET [184]. Using

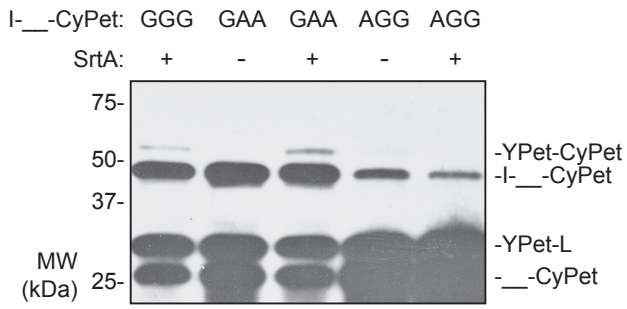

Figure 5.2: *In vivo* protein fusion of precurosrs with a single glycine. Western blot showing crosslinking of proteins that contain only one glycine after the intein. The cells in the SrtA (+) lanes have a SrtA-containing plasmid, while (-) have an empty plasmid. For I-__-CyPet, (GGG) is the intein followed by GGG residues, (GAA) is the intein followed by GAA residues, and (AGG) is the intein followed by AGG residues.

expression conditions that were optimized for fluorescent protein ligation, the FRET ratio for crosslinked YPet-CyPet was 1.25-fold higher than samples that lacked SrtA or contained a non-binding SrtA recognition sequence (**Figure 5.1c**). The FRET ratio is calculated by normalizing the FRET signal (ex: 415, em: 527) by the total CyPet production (ex: 415, em: 478) to control for differences in precursor protein amount; CyPet is chosen for normalization since it is the donor in the FRET pair [186]. Although only a modest gain, it shows that it is possible to discern crosslinked proteins *in vivo* using FRET. It is believed that the low amount of fused precursor proteins and poor FRET signal in bacteria due to auto-fluorescence are the cause for not observing a larger increase in signal above background.

SrtA crosslinking can be generalized to other recombinant proteins. To test the robustness of the SrtA fusion technique, the YPet and CyPet proteins were replaced with other soluble *E. coli* proteins such as: maltose-binding protein (MBP), glutathione s-

transferase (GST), β-lactamase (Bla), green fluorescent protein (GFP), enolase (Eno) [10], and scFvs. Using optimized expression conditions, both the YPet and CyPet could be substituted and desired fusions are created *in vivo* (**Figure 5.3a-c**). As shown previously, these fusions only occurred when SrtA was co-expressed and the proper set of tags were included. The ligation efficiency was lower than that observed for the YPet and CyPet fluorescent proteins. Due to the lower production of certain recombinant proteins, the small amount of fusion, and sensitivity of antibody detection, all the protein-protein fusions could not be observed inside cells. Rather, the precursor proteins were produced separately and lysates mixed to allow ligation to occur (**Figure 5.3d-f**). Eliminating co-expression increased the amount of protein produced and longer incubations with SrtA allowed for more ligations as to be visible by antibody detection on western blots. Ligation *in vitro* was dependent on both the presence of SrtA and the GGG-recognition sequence. A goal of this work is to create a tandem scFv by post-translational fusion due to their potential as novel therapeutics and low soluble production in *E. coli* [153-155]. Here we crosslinked an scFv targeted for bacteriophage capsid protein D known as $scFv_{gpD}$ [187] with one that binds beta-galactasidase called $scFv_{R4}$ [188] using SrtA (**Figure 5.3f**). Although, the amount of tandem scFv produced by enzymatic fusion is low, it is possible to create a soluble and covalently linked scFvs post-translationally, eliminating refolding the scFv fusions from inclusion bodies.

***In vitro* characterization of SrtA.** To prove the protein-protein crosslinking observed in *E. coli* can occur between two folded and active proteins, we carried out *in vitro* fusion

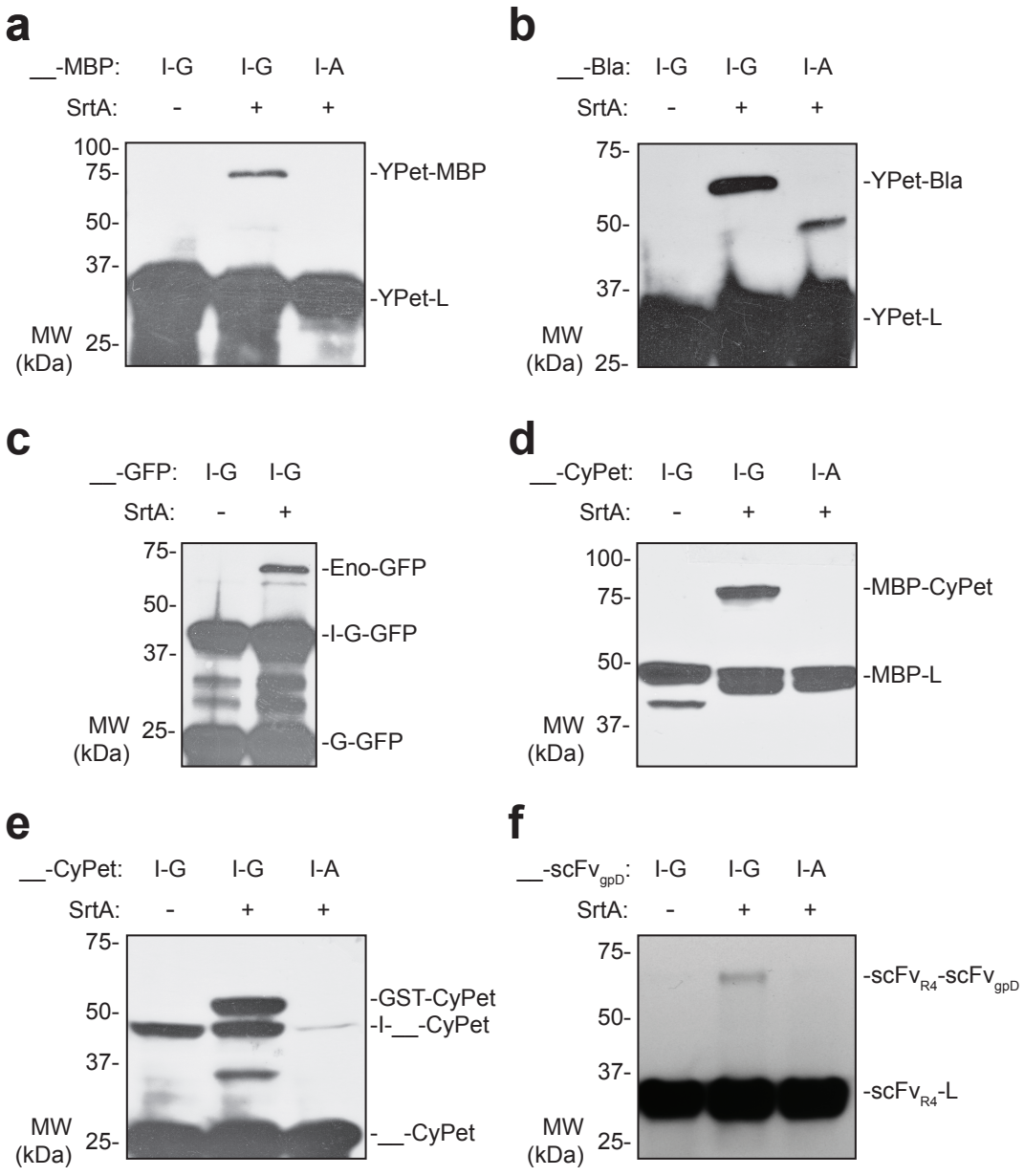

Figure 5.3: Robust post-translational fusion of recombinant proteins. A-C) Western blots showing the crosslinking of several recombinant proteins A) YPet to MBP, B) YPet to Bla, and C) Eno to GFP in E. coli. The cells in the SrtA (+) lanes have a SrtA-containing plasmid, while (-) have an empty plasmid. D-F) Western blots showing the fusion of proteins in vitro using lysates that contained separately produced recombinant proteins with fusion tags; D) MBP to CyPet, E) GST to CyPet, and F) scFvR4 to scFvgpD. Samples were treated with (+) or without (-) SrtA-containing lysate. (I-G) proteins are preceded by an intein followed by GGG residues, while (I-A) samples have an intein and AAA residues. Proteins were detected using anti-GFP (A-C,E), anti-MBP (D) or anti-cmyc (F) antibodies.

experiments using purified YPet and CyPet proteins. As was found for *in vivo* crosslinking of YPet and CyPet, ligation only occurred when SrtA-recognition tags were included with SrtA (**Figure 5.4a**). When the non-specific tags were included on CyPet and YPet, fusion was not observed. Crosslinking did not occur when one of the precursor proteins was eliminated, suggesting no undesired fusion products were created. The majority of the intein was cleaved from the Intein-__-CyPet during the purification and storage of CyPet. However, the crosslinking reaction yielded far from complete YPet to CyPet ligation with excess precursor proteins still available, which may suggest the reaction is kinetically limited due to weak binding of substrate. A more likely possibility is that after the cleavage of the G-his tag from YPet-LPETG-his by SrtA, this substrate will compete with the GGG-CyPet for ligation to YPet-LPET. Since both of these molecules contain N-terminal glycines, they are equally recognized by SrtA and will result in an equilibrium between the YPet-LPETGGG-CyPet fusion product and reformed YPet-LPETG-his species. Due to these substrates being present in close to equal molar amounts after the initial cleavage of the G-his (**Figure 5.4a**), the regeneration of YPet-LPETG-his will limit the amount of desired product that can be successfully formed.

In addition to characterizing crosslinking by measuring end point creation of protein fusions, we directly observed crosslinking in real-time using FRET (**Figure 5.4b**). As expected, when SrtA is removed from the reaction or either non-specific recognition tag is used, the FRET signal is unchanged throughout the reaction. Upon addition of SrtA, an increase in FRET ratio is observed as crosslinking occurs. The rate

of FRET increase is dependent on the concentration of SrtA (**Figure 5.4b**), YPet-LPETG, and Intein-GGG-CyPet (**Figure 5.5**). Using this data we had hoped to determine the kinetic rate constants for protein substrates and SrtA; however, due to the high K_m for the LPETG substrate reaction, we were unable to saturate the reaction to accurately determine Michaelis-Menten constants. Based on these experiments, we confirm the K_m for an LPETG-containing protein in the presence of saturating amounts of Ca^{+2} to be greater than 50 μM but it is unclear if it is in the high micromolar or low millimolar range [175, 181]. It is clear that binding to LPETG is rate limiting as with 50 μM YPet-LPETG, a K_m for GGG-CyPet of 4.1 μM was found. However, equilibrium amounts of YPet-LPETG-his and YPet-LPETGGG-CyPet will be formed as mentioned previously that will depend on the initial amounts of YPet-LPETG and GGG-CyPet. Both the intial cleavage of the G-his from YPet-LPETG-his and formation of ligation product must be considered simultaneously to determine kinetic constants using this method. To our knowledge, this is the first attempt to measure SrtA crosslinking rates *in vitro* using protein substrates and in real-time.

Intein removal results in lower crosslinking. To test the hypothesis that an N-terminal glycine residue is necessary for crosslinking, an MGGG-CyPet protein was created that lacked the intein domain. It was thought that ligation would be blocked by the N-terminal start codon (Met_f). However, when this protein was co-expressed *in*

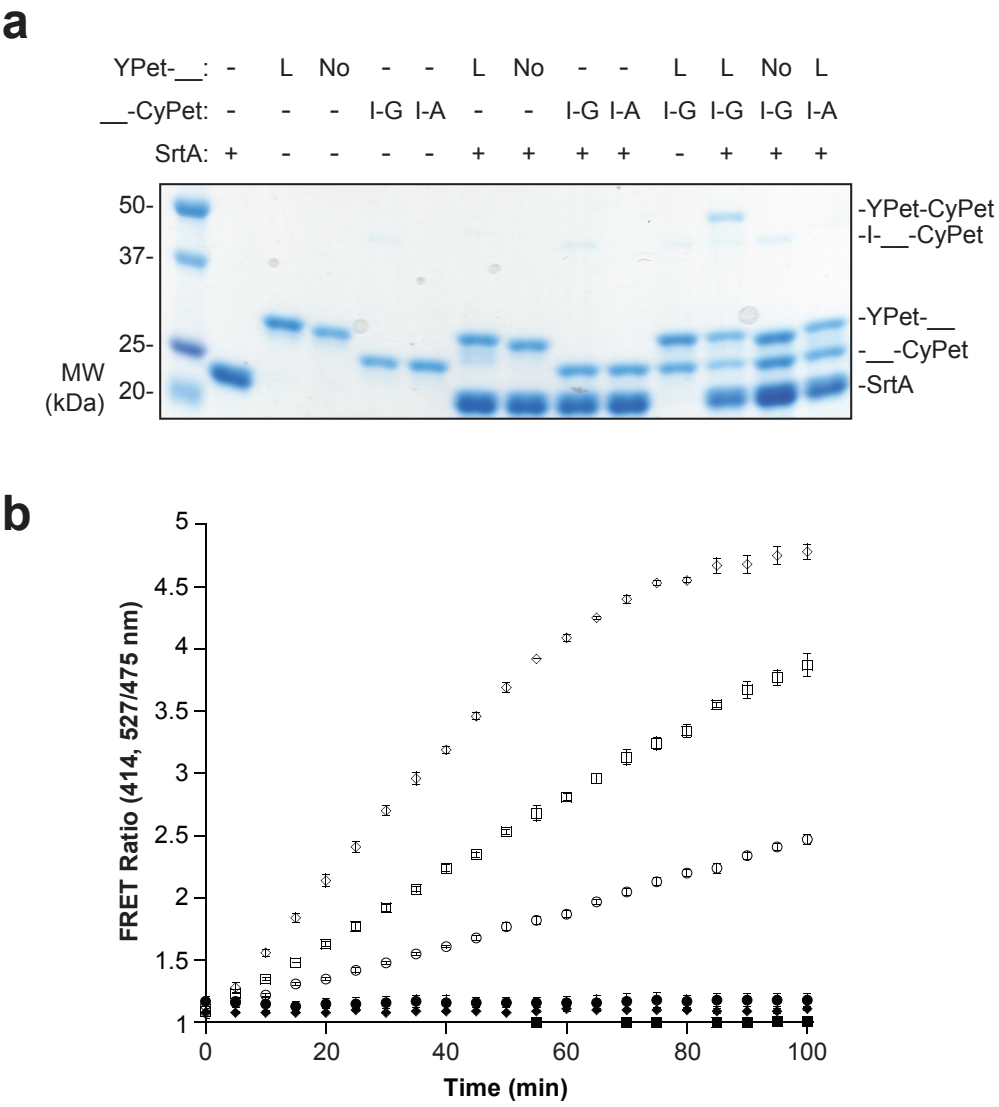

Figure 5.4: *In vitro* crosslinking of purified fluorescent proteins. A) *In vitro* crosslinking reaction that was performed for 21 hours at 37ºC and resulting Coomassie-stained gel. YPet-__: (-) not present, (L) contain the LPETG tag, and (No) have a GGGSG tag. __-CyPet: (-) not present, (I-G) contain the intein-GGG tag, and (I-A) have an intein-AAA tag. SrtA: (-) not present and (+) with SrtA. B) Using FRET to monitor the SrtA crosslinking reaction. Samples are as follows: closed circle (10 μM YPet-LPETG, 10 μM Intein-GGG-CyPet and 0 μM SrtA), closed square (10 μM YPet-LPETG, 10 μM Intein-AAA-CyPet and 25 μM SrtA), closed diamond (10 μM YPet-GGGSG, 10 μM Intein-GGG-CyPet and 25 μM SrtA), diamond (10 μM YPet-LPETG, 10 μM Intein-GGG-CyPet and 50 μM SrtA), square (10 μM YPet-LPETG, 10 μM Intein-GGG-CyPet and 25 μM SrtA), and circle (10 μM YPet-LPETG, 10 μM Intein-GGG-CyPet and 10 μM SrtA). FRET ratio is the FRET signal (Ex: 430 nm, Em: 527) divided by the CyPet signal (Ex: 430 nm, Em: 527); error bars represent the standard deviation of 3 replicates.

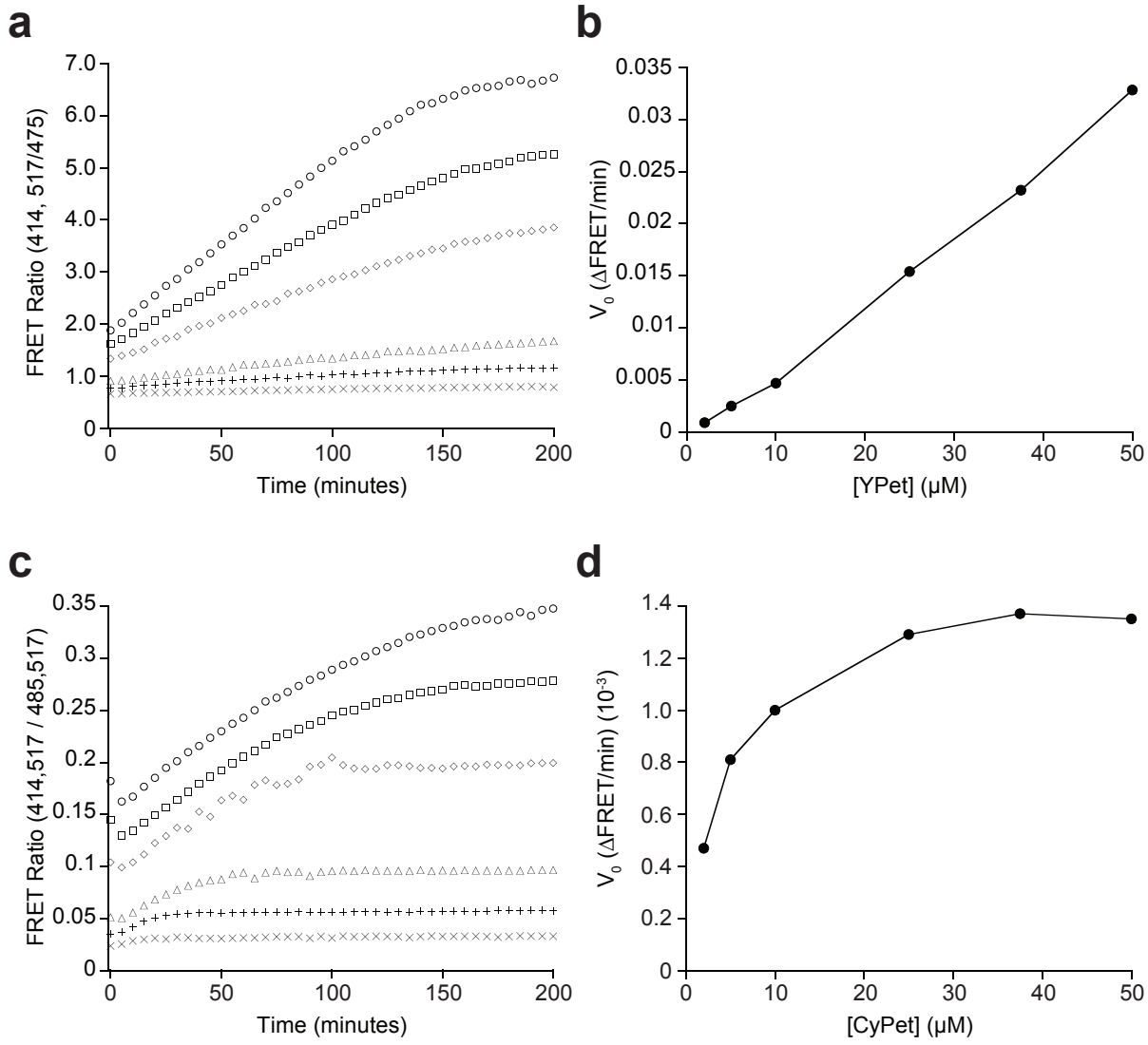

Figure 5.5: SrtA kinetics measured *In vitro* using FRET. A) *In vitro* protein ligation with 50 μM Intein-GGG-Cy-Pet, 25 μM SrtA, and various concentrations of YPet-LPETG. YPet-LPETG concentrations are as follows: circle (50 μM), square (37.5 μM), diamond (25 μM), open triangles (10 μM), plus (5 μM), and cross (2 μM). FRET ratio is the same as defined previously. B) Initial rates of ligation (ΔFRET/time) as a function of YPet-LPETG concentration. C) *In vitro* protein ligation with 50 μM YPet-LPETG, 25 μM SrtA, and various concentrations of GGG-CyPet. GGG-CyPet concentrations are as follows: circle (50 μM), square (37.5 μM), diamond (25 μM), triangle (10 μM), plus (5 μM), and cross (2 μM). FRET ratio is the FRET signal (ex: 414 nm, em: 517) divided by YPet signal (ex: 485, em: 517) since YPet concentration was held constant for this experiment. D) Initial rates of ligation (ΔFRET/-time) as a function of GGG-CyPet concentration.

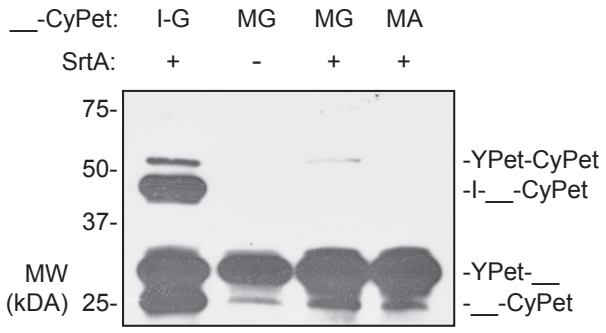

Figure 5.6: Protein fusion *in vivo* without an intein. Comparison of protein cross-linking inside E. coli after changing the N-terminal tag on CyPet. (I-G) represents the Intein-GGG-CyPet protein, (MG) samples lack the intein and the CyPet is preceded by a start codon and GGG sequence. The (MA) samples also lack the intein, and a start codon and AAA residues precede CyPet. The cells in the SrtA (+) lanes have a SrtA-containing plasmid, while (-) have an empty plasmid.

vivo with YPet-LPETG, a protein-protein ligation product was detected (**Figure 5.6**). It is known that natural *E. coli* methionine aminopeptidase removes the Metƒ [189-191], which would expose an N-teminal glycine and allow SrtA to crosslink proteins. When the MGGG tag is replaced by MAAA on CyPet, protein fusion is abolished. Additionally, when SrtA is not coproduced, no fusion band is detected. The removal of the intein tag results in a decreased amount of overall crosslinking. It is believed that there is less precursor GGG-CyPet due to the lack of intein stabilization, resulting in lower overall fusion.

Discussion

Here we report a simple and robust method for enzymatic ligation of proteins inside *E. coli.* Protein fusion amount and efficiency is largely influenced by expression and SrtA kinetics. The K_m for SrtA for its LPXTG recognition sequence has been reported in the high micromolar range, which is above the concentration of many

intracellularly produced proteins. Additionally, the recognition of the LPXTG peptide by SrtA is influenced by Ca^{+2} with an affinity of around 2 mM [178]. Intracellular Ca^{+2} concentrations in *E. coli* are around 90 μM [192] and much below the K_d for SrtA, resulting in decreased affinity for its recognition tag. In addition to the Ca^{+2} requirement and high K_m, the catalytic cysteine of SrtA must be present in the correct ionized form, which may not be possible under physiological conditions [159]. Intein cleavage is not complete *in vivo*, limiting the SrtA substrate pool and subsequent crosslinking. However, expression of a protein containing an MGGG tag resulted in less protein-protein ligation than when the intein was included due to lower precursor production and incomplete processing of the Met_f. Despite these limitations, we show for the first time protein-protein crosslinking in the cytoplasm of *E. coli* for several recombinant proteins.

Furthermore, we have developed a fluorescent protein based assay, which has shown that SrtA reaction with an LPXTG-containing protein has a lower K_m than the ligation reaction to a protein with an N-terminal glycine. However, the G-his peptide product that results from the initial SrtA cleavage competes with the GGG-CyPet for fusion to YPet-LPET. Excess GGG-CyPet should shift the equilibrium towards the desired fusion product. Unfortunately, the high K_m value for the LPETG-containing peptide recognition kinetically limits this enzyme for protein-protein ligation since it is not possible to add GGG-CyPet in mM concentrations. Even though FRET is readily measured in the *in vitro* ligation reaction, it is only detectable slightly above controls inside cells. This is due to the limited crosslinking efficiency as well as a poor signal to noise ratio for FRET in bacteria due to autofluorescence. Unfortunately, this restricts

the use of FRET as a screening tool for developing evolved SrtAs with better kinetics. Even though directed evolution strategies exist for SrtA, they have yet to produce variants with increased kinetics for both SrtA reaction steps [181]. Designing genetic selections by challenging with competitive LPETG-containing substrates could be useful in designing an SrtA with enhanced properties. In order for enzymatic crosslinking to be useful in post-translational fusion strategy, more efficiency is necessary, requiring an SrtA or other enzyme that has reaction kinetics better suited for intracellular conditions.

Materials and methods

Bacterial strains and plasmids. MC4100 cells were used for cloning and the express proteins from pBAD18 and pTRC99(cm). BL21(DE3) cells were used to express proteins from pOKD4, pET21a and pET28a. The plasmid pBAD18 was modified for easy cloning of protein-ligation targets as well as to facilitate the combining of plasmids for co-expression and subsequent *in vivo* fusion (**Figure 5.7**). Protein-ligation targets were cloned between the FseI and KpnI cut sites in pJTB or pBTJ. The protein containing the SrtA recognition peptide was made in pJTB, while the intein-GGG protein was made in pBTJ. Genes from pBTJ could be cloned polycistronically in pJTB by using the AvrII and SphI cut sites (**Figure 5.7**). The *srtA* gene was cloned singularly in pTRC99(cm) or pOKD4 for use in *in vivo* crosslinking. The plasmid pOKD4 is further described in [193].

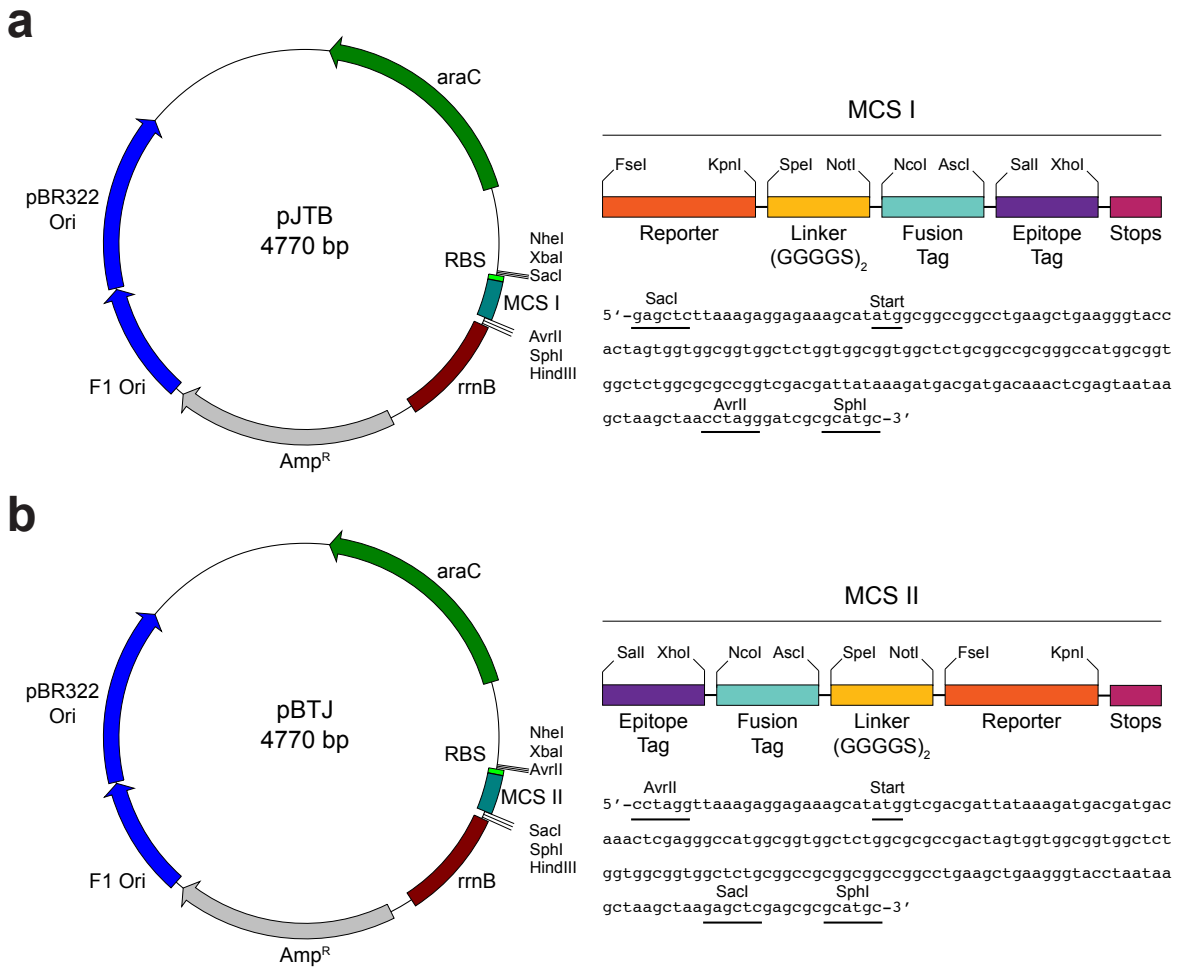

Figure 5.7: Reporter plasmid constructs for *in vivo* protein ligation. Both the pJTB and pBTJ are based on pBAD18. MCS I A) was combined with MCS II B) in the pJTB plasmid using the AvrII and SphI restriction enzyme sites to create a polycistronic plasmid that was used to express both precursors for *in vivo* protein-protein ligation. Each MCS is presented to show the modularity of the plasmids.

***In vivo* protein-protein crosslinking.** MC4100 cells were freshly co-transformed with pBAD18 that contained the genes of the two proteins to be crosslinked with desired tags, and with pTRC99(cm) SrtA. Additionally, pTRC99(cm) (empty) was also transformed pBAD18 as a negative control. All transformants were plated on LB agar containing appropriate antibiotics. Colonies were selected and grown overnight at 37°C

in LB and appropriate antibiotics. Cells were subcultured into shake flasks with LB and antibiotics. Cultures were grown at 37°C for 2 hours until an OD_{600} of approximately 0.6 was reached. Protein production was induced for using 0.1 mM IPTG and 0.2% arabinose. Protein expression occurred at 22°C for 12 hours, followed by 12 hours at 37°C. Other expression temperatures and lengths of time were used depending on the proteins to be crosslinked; the two-temperature induction was determined to be optimal for YPet and CyPet fusion. After induction, the OD_{600} of each culture was measured, and cultures were normalized by adding LB. Whole cells were lysed by boiling with an equal volume of 2xSDS buffer for 15 minutes. Proteins were separated by SDS-PAGE and were transferred to a PVDF membrane for western blotting. For YPet, CyPet and GFP, blots were probed using a 1:15000 dilution of anti-GFP (Roche) and 1:2500 dilution of HRP-tagged anti-mouse (Pierce).

FRET Measurements *in* vivo. BL21(DE3) cells were freshly transformed with pET21a YPet-LPETG-his, Intein-__-CyPet-his and pOKD4 SrtA. The pOKD4 (empty) plasmid was also transformed with the pET21a reporter plasmid as a negative control. Three colonies from each plate were picked and grown overnight at 37°C in LB and appropriate antibiotics. Cells were then subcultured into shake flasks with LB and antibiotics to an OD_{600} of 0.05. Cultures were grown at 37°C for 2 hours until an OD_{600} of approximately 0.6 was reached. Protein production was induced for using 0.1 mM IPTG and cultures were moved to 22°C for 8 hours. Following the initial incubation for fluorescent protein production, cultures were placed at 37°C and time points were taken

at 0 and 26 hours. OD_{600} measurements were made and cells were harvested by centrifugation; 1 mL of cells were pelleted for an OD_{600} = 2.5. Cells were then resuspended in 200μL of 1xPBS and fluorescent measurements were made on a Gemini II 96-well plate reader (Molecular Devices) using black-lined 96-well plates. The amount of CyPet (blue fluorescence) was measured by exciting the sample at 414 nm collecting emission at 475 nm. The amount of YPet (yellow fluorescence) was determined using an excitation of 485 nm and emission of 517 nm. For FRET, CyPet serves as the donor molecule and YPet the acceptor due to their spectral overlap. To measure the FRET signal, an excitation of 414 nm and emission of 517 nm was used. The cutoff filter for both emission was set based on the emission wavelength. The FRET signal for each sample was divided by its CyPet signal to give a normalized FRET ratio (414, 517/475).

Protein fusion of separately produced proteins. All pBAD18 and pTRC99(cm) plasmids were freshly transformed into MC4100 cells. Overnight cultures were subcultured to an OD_{600} of 0.05 and grown for 2 hours at 37°C. Protein expression was induced with 0.2% arabinose for pBAD18-containing cultures and 0.1 mM IPTG for pTRC99(cm)-containing cultures. Protein production was carried out 37°C, 30°C, 22°C for 24 to 30 hours. Different expression temperatures were used for each recombinant protein depending on which gave rise to the highest amount of protein; SrtA was produced at 37°C for 3 hours. Cells were pelleted by centrifugation at 4,000xg for 10 minutes. Cell pellets were resuspended in 1xPBS and lysed by sonication. Soluble

protein was recovered by centrifugation at 16,000xg for 30 minutes. SrtA crosslinking reactions were set up to contain an equal volume of SrtA-containing lysate, LPETG-tagged protein lysate, and Intein-tagged protein lysate. Crosslinking reactions were carried out at 37°C for 24 hours. Protein fusion was detected by western blot of crosslinking reactions. The following primary antibodies were used to detect proteins: anti-MBP (NEB) for MBP, anti-GFP (Roche) for CyPet, anti-cmyc (Sigma) for $scFv_{R4}$, and anti-HA (Sigma) for $scFv_{gpD}$.

***In vitro* fluorescent protein fusion.** The YPet and CyPet proteins were produced from a pBAD18 vector using a 0.2% arabinose induction for 18 hours at 22°C. SrtA was expressed from pET28a using 1mM IPTG induction for 3 hours at 37°C. Cell pellets were collected by centrifuging cultures at 4,000xg for 10 minutes. Cells were resuspended in 20 mM Tris HCl, 0.5 M NaCl, 5 mM imidazole (pH 7.9) and lysed by sonication. Soluble protein was recovered by centrifugation at 30,000xg for 30 minutes. Proteins were purified using HisBind resin (Novagen) and gravity-based protocol according to the manufacturer's recommendations. Eluted proteins were dialyzed into 50 mM Tris (pH8), 150 mM NaCl by three buffer exchanges, and were concentrated using a 10,000 kDa molecular weight cutoff centrifuge filter (Millipore). Proteins were stored in 10% glycerol at 4°C. Protein concentration was quantified by Bradford assay (BioRad) using a BSA standard curve.

Purified component crosslinking experiments were performed using 10 μM each YPet and CyPet precursors in the presence and absence of sortase (25 μM) in 50 mM

Tris (pH8), 150 mM NaCl, 10 mM $CaCl_2$ and 10% v/v glycerol at 37°C for 21 hours. Proteins were separated by SDS-PAGE and the resulting gel was stained with Coomassie blue to detect all proteins species present. Image analysis was performed using ImageJ. For *in vitro* FRET measurements, samples were made with desired concentrations of SrtA, YPet and CyPet in in 50 mM Tris (pH8), 150 mM NaCl, 10 mM $CaCl_2$ and 10% v/v glycerol. Samples were made to 250 µL in a black-lined 96-well plate. YPet (ex: 485 nm, em: 517), CyPet (ex: 414, em: 475), and FRET (ex: 414, em: 517) fluorescence measurements were recorded every 5 minutes for a period of 100 minutes using a Gemini II 96-well plate reader (Molecular Devices). The cutoff filter for the emission was set based on the emission wavelength. The reaction-containing plate was incubated at 37°C for the duration of the reaction and was mixed every 30 seconds.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

In this work, we used an authentic bacterial quality control mechanism to enhance the production of a recombinant endocellulase 30-fold after just two directed evolution cycles. Expression increase was due to just two mutations in the catalytic domain of the cellulase and occurred without activity changes on soluble or insoluble cellulose substrates. This work represents the first time Tat-based protein engineering was employed for protein production enhancement while maintaining high enzyme activity, due to the two-tiered selection and screening method [56, 59]. Further, it was found that the evolved cellulase had a decreased thermal stability and folding free energy compared to the parent protein; however, it is currently unknown whether those changes can be generalized to other proteins selected using the Tat-based scheme. The cellulase was shown to have a similar “tadpole” shape that is characteristic of endocellulases [143, 144], suggesting this property was conserved through the directed evolution process. We have begun probing how selective pressure can influence the selectivity of the quality control-based engineering technique, discovering that high selective pressures biases fitness improvements towards solubility enhancement at the cost of functional loss. Lastly, our desire to determine the applicability of Tat quality control towards other protein targets led us to study the TatABC translocon itself,

discovering the potential chaperone-like role of TatB and its newly defined role in translocation.

Future directions

Quality control-based approaches for protein engineering. As discussed in the introduction, utilizing cellular quality control mechanisms to engineer protein folding has great potential to influence heterologous protein production and narrowing protein libraries [42]. In nature, mutated proteins are subject to quality control analysis, removing unfit proteins and thus organisms from the population. The Tat-based folding reporter has been successful at selecting members with improved production [55, 56, 59, 126]; however, questions still remain as to the selectivity of the reporter, what biophysical properties are altered to give rise to an expression increase, and how robust is the approach. We began studying the selectivity of the Tat-based reporter by comparing percentage of active members at different selective pressures. Application of the selection step seemed to only modestly focus the sequence space towards active members. In the future, more library members must be compared to determine if the selection step is having the desired effect of narrowing the libraries. Additionally, isolated members from each selective pressure should be characterized (e.g. solubility, mutations, specific activity, stability) to ensure the selection is working as expected. We plan to use these experiments better understand function vs stability tradeoffs [103] as well as to begin building a fitness landscape to describe the quality control-based engineering approach, selectivity/solubility vs active members [194].

Determining the cause of enzyme fitness improvements was a focus of this work and will influence how the approach is used in the future. For Cel5A TD, production enhancement was likely due to a general cytoplasmic solubility increase rather than specific engagement of Tat quality control. When the cellulase was expressed without the Tat-targeting signal sequence or Bla, the expression increase was maintained. This suggests that the isolated protein fusion is overproduced in the cytoplasm, leading to increased translocation and a selective growth phenotype. In addition to the biophysical characteristics tested in Chapter 2, expression enhancement can be due to differential interactions with cytoplasmic chaperones, elimination of kinetically-trapped folding intermediates, alteration of the cellulase structure to be more “bacteria-like”, or abolition of the need for disulfide bonding during the folding process. The catalytic domain of the Cel5A wt and evolved Cel5A TD could be crystalized to determine the structural changes occurred and to possibly narrow down what characteristics should be probed in the future. Additionally, other selected library members need to be characterized to see how generalizable the biophysical alterations are. Lastly, the Tat-based reporter should be compared to other cytoplasmic reporters that use GFP (whole or split) [46, 52], DHFR [47], or Cat [48] to directly determine if an advantage is provided by including the translocation quality control step. Although, engineering heterologous protein production has been successful, it is time consuming and so far has not yielded general rules for expression enhancement [6]. In the future, such methods should be used sparingly for difficult to produce cytoplasmic proteins, while general strain improvements

[8, 12] (e.g. disulfide bonding, protease deletion) or chaperone engineering [121] maybe employed more globally.

Protein engineering meets metabolic engineering. Since the first mention of metabolic engineering in the early '90s [195, 196], the field has greatly progressed over the past 25 years due to advancements in genomic sequencing, enzyme characterization and pathway implementation [197]. Today, databases such as Kegg allow scientists to select enzymes from diverse organisms to integrate chemistries of interest into platform organisms [198]. Metabolic engineers have created a vast array of chemicals from second-generation biofuels [199-203] to secondary metabolites such as terpeniods, polyketides and non-ribosomal peptides [204-206]. One of the largest success stories of metabolic products, due to the scale of its production, is 1,3-propanediol being produced in *E. coli* by Dupont and used as a polymer called Sorona in carpets [207]. Biofuels such as isobutanol [208] and alkanes [209] can now be made in bacteria through incorporation of just two heterologous enzymes. Additionally, high value therapeutic precursors are being created, such as artemisinic acid that can be converted into antimalarial drugs [210]. Diversion of natural pathways combined with incorporations of specific enzyme chemistries aims to rapidly expand the range of bioproducts being formed [198].

In 2013, the Defense Advanced Research Projects Agency released a call for their Living Foundries program. The goal of the Living Foundries is to create centers to combine research in DNA synthesis, expression tuning, enzyme pathway selection, and

strain engineering to determine fundamental design principles to expand the scalable production of biological molecules. As a proof of principal, they set out to create 1000 unique compounds in ten foundries. This program represents a shift in synthetic biology research from “parts” generation for a larger biological “toolbox”, to implementing the pieces of pathways together for novel compound creation. However, the goal of integrating generalizable and interchangeable parts, the way a computer engineer designs a chip, is far from realization and may be impossible due to the complexities of biological systems, our incomplete understanding of protein mechanisms, and the inherent robustness of cells to resist change from their evolved purpose [197, 211]. In a recent review, several leading synthetic biologist describe many of design elements needed for successful pathway integration along an axis of understanding (describe-explain-predict-control); expression control and DNA manipulation are some of the most understood parameters, while protein folding, multi-protein interaction, allosteric regulation, translocation and eukaryotic expression are far from being controlled [211].

Enzyme engineering is a powerful technique that will greatly affect the success of heterologous metabolic pathway incorporation [105, 212]. Protein modification can alter expression amounts, kinetics (including selectivity, speed, and reversibility), cellular localization, toxicity, protein-protein interaction, and enzyme regulation, which are all critical design principals present in natural pathways [202, 213]. However, to date most metabolic engineering attempts have used traditional nucleic acid engineering [214, 215], multi-gene inserts [216], homologous enzymes [209], or genomic knockouts [217] to solve flux balance problems. A few examples of protein engineering influencing

metabolic production are highlighted below. Enzyme engineering of geranylgeranyl diphosphate synthase and levopimaradiene synthase has been used to increase production of a diterpenoid 2,600-fold by altering enzyme kinetics and promiscuity [218]. Clever use of yeast surface display enabled the discovery of the non-ribosomal peptide synthase with over 200-fold increase in selectivity, allowing it to be used to create specific peptides rather than heterogeneous products [219]. Cytochrome P450s have served as a flexible scaffold to permit the enzyme to perform chemistries not known in nature with greatly altered substrate profiles [220]. Lastly, the work here has used a folding reporter to develop heterologous enzymes with specific and selectable production amounts inside cells, which would not be possible using conventional nucleic acid engineering due to the low solubility of the parent enzyme.

All of this work will contribute to the increasing use of protein engineering to aid in designing heterologous pathways. Although, protein engineering is far from having design rules to predict alterations for specific fitness gains, its success towards a variety of applications poises it for use in metabolic engineering. Similarly, synthetic biology aims to build pathways from the ground up through detailed understanding and control over each pathway element [211]. Metabolic engineering has taken a top-down approach to generate biomolecules by employing enzymes and whole pathways without optimization from a variety of organisms. The lack of high-throughput screens has necessitated this top-down methodology due to the limitation of combinations that could be feasibly tested. Successes combining protein engineering with metabolic engineering have hinged on detection of a colorimetric product such as lycopene [218].

Developing clever screens using surface display [219], binning techniques [221], or *in vivo* sensors will be critical to optimizing pathway elements and further elucidation of rules that constrain design elements [211].

Next generation enzyme targets for the Tat-based genetic selection. Based on the discussion herein, protein engineering is primed to play a key role in the production of metabolic products in the near future. Heterologous enzyme expression is the first hurdle that must be crossed to endow cells with desired chemistries, and opens up the possibility of further enzyme engineering for increased reaction rates and altered specificity. Three key design principals must be considered when choosing the next enzyme candidate for production enhancement via the Tat-based genetic selection. First, enzymes that give rise to high value products should be given priority over those to make commodity chemicals. To paraphrase a message from Jay Keasling, it is imperative to show how the properties of biologically-derived molecules surpass those synthesized by traditional chemistry techniques to prove the potential of these high value products rather than competing with existing processes for minor gains [222]. Second, it is desired to have a low-throughput, expensive, and/or time-consuming screen to couple with the genetic selection. If the quality control based selection does in fact focus libraries towards folded and active members, coupling the selection with a "high-cost" screen will enable a higher likelihood of isolating members that will enhance titers and lower the amount of unfolded members to be screened. Lastly, it is important to consider enzymes whose production enhancement is desired in the soluble

cytoplasm of *E. coli*. Other reporters for periplasmic folding should be considered for enzymes targeted outside of the cytoplasm [49] and non-secretion reporters, such as Cat [48], DHFR [47] or split proteins [52], could be used for membrane bound enzymes. One promising candidate already studied by the DeLisa laboratory is DkgA [223], which is the central enzyme in a heterologous pathway for 1,2 propanediol production and whose expression has not been detected.

In an excellent review, Tang and co-workers describe recent advances in enzyme engineering for metabolic product generation [212], highlighting classes of enzymes that could be included in the Tat-based genetic selection. From that work, it was found that the most promising candidates for *E. coli* quality control based engineering are polyketide synthases. Polyketides are a class of large, complex, and chiral compounds polymerized from simple acyl-coenzymeAs precursors, and have wide application as antibiotics, cancer therapeutics, or precursors for commodity chemicals [224-226]. Polyketide synthases are naturally produced by Actinomycetes, Myxobacteria, and filamentous fungi and have proven difficult to express in *E. coli* even though the requisite post-translational modifications appear to be present [226, 227]. The enzymes are modular in function, forming pipelines to create molecules such as triketide lactones [228] and epithiolone [229]. Due to the complex and chiral nature of the compounds produced as well as their low titers, screening for successful product formation by liquid chromatography or gas chromatography coupled with mass spectrometry is incredibly low throughput, costly, and time consuming [226]. All of these reasons make them the ideal enzymes for the Tat-based selection. Due to their

modularity, successful expression of an enzyme can lead to generation of multiple products through combination with other polyketide synthases [225, 227]. Additionally, production enhancement design principals can be applied to other polyketide synthases to further narrow the sequence space to be searched [226]. The Tat-based folding reporter for soluble enhancement could possibly be combined with a protein fragmentation reporter to increase enzyme-enzyme complex formation in *E. coli*, which is critical to substrate channeling in polyketide synthases [225] and proven difficult in bacteria [9].

Heterologous hosts beyond *E. coli*. Although *E. coli* is still the preferred host for heterologous protein production, numerous other organisms aim to or already out compete it in the near future for specific applications. For instance, due to their high culture titer and ability to secrete glycosylated proteins, Chinese Hamster Ovary cells are the most widely used host for monoclonal antibody production [3]. In addition to pathway advantages provided by specific organisms, the understanding and manipulation of biological systems has greatly expanded over the past 15 years. The well-understood and easy manipulation of *E. coli* genetics not only enabled it as platform organism, but also paved the way for further study and refinement of these mechanisms for use in other hosts. Genomic editing tools are becoming commonplace in laboratories and are now available for eukaryotes [230]. DNA synthesis and library generation through microarrays is progressively less expensive, eliminating the need for typical cloning techniques [211, 231]. Genomic sequencing of organisms has compiled

laundry lists of heterologous protein candidates as well as enabled the discovery of novel pathways. The "-omics revolution" is profiling protein and metabolite pools, which can greatly impact the ease of recombinant protein incorporation. Lastly, bioreactor design has permitted the culturing of organisms to high densities that was previously not possible.

For all these reasons, the advantages of using *E. coli* as an expression platform can be substituted by a number of other platform organisms. As evidenced by antibody production in mammalian cells, the presence of secretion and glycosylation pathways provided reasons to switch hosts. In addition to post-translational modification pathways, the presence of other enzyme pathways to create novel secondary metabolites or alter the energetics of specific cellular compartments is a main concern for metabolic engineering in the future. For instance, proteorhodopsins have been produced in *Shewanella* that also contains an electric transport chain to allow the organism to generate electricity and increase metabolism [232]. Photosynthetic hosts or ones that grow on industrial waste streams offer the advantage of having even cheaper media formulations than defined media for *E. coli* growth. Cyanobacteria are becoming increasingly popular as a host organism due to their growth on carbon dioxide and sunlight [233]. Recently, they were engineered to produce isobutyraldehyde, which can be converted into other products including fuels [234]. In addition to growing on practically free resources, cyanobacteria offer advantages over cellulose conversion by eliminating the energy intensive (and wasteful) polymerization and deconstruction steps.

Using natural pathway architectures that have been evolved for thousands or chemistries specific to organisms of interest can be far simpler than recreating the entire system in *E. coli*. However, even with our advanced knowledge of biological systems and ability to characterize them, movement away from this model organism or yeast will complicate heterologous protein expression. Design of folding reporter selections or screens in these organisms will aid in the transfer of pathways into new hosts by studying genetic regulation and engineering of specific proteins. Changing platform organisms will necessitate new research into what are controllable biological parts; however, the extensive research over the past 40 years should allow the new discoveries to occur quickly for heterologous protein production.

REFERENCES

[1] Lee SY, Mattanovich D, Villaverde A. Systems metabolic engineering, industrial biotechnology and microbial cell factories. Microbial cell factories. 2012;11:156.

[2] Johnson IS. Human insulin from recombinant DNA technology. Science. 1983;219:632-7.

[3] Aggarwal RS. What's fueling the biotech engine-2012 to 2013. Nature biotechnology. 2014;32:32-9.

[4] Xiao S, Shiloach J, Betenbaugh MJ. Engineering cells to improve protein expression. Current opinion in structural biology. 2014;26C:32-8.

[5] Baneyx F. Recombinant protein expression in Escherichia coli. Current opinion in biotechnology. 1999;10:411-21.

[6] Makino T, Skretas G, Georgiou G. Strain engineering for improved expression of recombinant proteins in bacteria. Microbial cell factories. 2011;10:32.

[7] Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied microbiology and biotechnology. 2006;72:211-22.

[8] Baneyx F, Mujacic M. Recombinant protein folding and misfolding in Escherichia coli. Nature biotechnology. 2004;22:1399-408.

[9] Netzer WJ, Hartl FU. Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature. 1997;388:343-9.

[10] Chang HC, Kaiser CM, Hartl FU, Barral JM. De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. Journal of molecular biology. 2005;353:397-409.

[11] de Marco A. Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in Escherichia coli. Nature protocols. 2007;2:2632-9.

[12] Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microbial cell factories. 2012;11:56.

[13] Schaerlaekens K, Lammertyn E, Geukens N, De Keersmaeker S, Anne J, Van Mellaert L. Comparison of the Sec and Tat secretion pathways for heterologous protein production by Streptomyces lividans. Journal of biotechnology. 2004;112:279-88.

[14] Choi JH, Lee SY. Secretory and extracellular production of recombinant proteins using Escherichia coli. Applied microbiology and biotechnology. 2004;64:625-35.

[15] Merlin M, Gecchele E, Capaldi S, Pezzotti M, Avesani L. Comparative evaluation of recombinant protein production in different biofactories: the green perspective. BioMed research international. 2014;2014:136419.

[16] Nevalainen KM, Te'o VS, Bergquist PL. Heterologous protein expression in filamentous fungi. Trends in biotechnology. 2005;23:468-74.

[17] Gimpel JA, Specht EA, Georgianna DR, Mayfield SP. Advances in microalgae engineering and synthetic biology applications for biofuel production. Current opinion in chemical biology. 2013;17:489-95.

[18] Natale P, Bruser T, Driessen AJ. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. Biochim Biophys Acta. 2008;1778:1735-56.

[19] Mansell TJ, Fisher AC, DeLisa MP. Engineering the protein folding landscape in gram-negative bacteria. Current protein & peptide science. 2008;9:138-49.

[20] Goemans C, Denoncin K, Collet JF. Folding mechanisms of periplasmic proteins. Biochimica et biophysica acta. 2014;1843:1517-28.

[21] Palmer T, Berks BC. The twin-arginine translocation (Tat) protein export pathway. Nature reviews Microbiology. 2012;10:483-96.

[22] Bogsch EG, Sargent F, Stanley NR, Berks BC, Robinson C, Palmer T. An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. The Journal of biological chemistry. 1998;273:18003-6.

[23] Palmer T, Sargent F, Berks BC. Export of complex cofactor-containing proteins by the bacterial Tat pathway. Trends in microbiology. 2005;13:175-80.

[24] Tullman-Ercek D, DeLisa MP, Kawarasaki Y, Iranpour P, Ribnicky B, Palmer T, et al. Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides. The Journal of biological chemistry. 2007;282:8309-16.

[25] Lee PA, Tullman-Ercek D, Georgiou G. The bacterial twin-arginine translocation pathway. Annual review of microbiology. 2006;60:373-95.

[26] Fisher AC, DeLisa MP. Laboratory evolution of fast-folding green fluorescent protein using secretory pathway quality control. PloS one. 2008;3:e2351.

[27] Graubner W, Schierhorn A, Bruser T. DnaK plays a pivotal role in Tat targeting of CueO and functions beside SlyD as a general Tat signal binding chaperone. The Journal of biological chemistry. 2007;282:7116-24.

[28] Ramasamy SK, Clemons WM, Jr. Structure of the twin-arginine signal-binding protein DmsD from Escherichia coli. Acta crystallographica Section F, Structural biology and crystallization communications. 2009;65:746-50.

[29] Pommier J, Mejean V, Giordano G, Iobbi-Nivol C. TorD, a cytoplasmic chaperone that interacts with the unfolded trimethylamine N-oxide reductase enzyme (TorA) in Escherichia coli. The Journal of biological chemistry. 1998;273:16615-20.

[30] Tarry MJ, Schafer E, Chen S, Buchanan G, Greene NP, Lea SM, et al. Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. Proceedings of the National Academy of Sciences of the United States of America. 2009;106:13284-9.

[31] Maurer C, Panahandeh S, Jungkamp AC, Moser M, Muller M. TatB functions as an oligomeric binding site for folded Tat precursor proteins. Molecular biology of the cell. 2010;21:4151-61.

[32] Panahandeh S, Maurer C, Moser M, DeLisa MP, Muller M. Following the path of a twin-arginine precursor along the TatABC translocase of Escherichia coli. The Journal of biological chemistry. 2008;283:33267-75.

[33] Frobel J, Rose P, Lausberg F, Blummel AS, Freudl R, Muller M. Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB. Nature communications. 2012;3:1311.

[34] DeLisa MP, Tullman D, Georgiou G. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proceedings of the National Academy of Sciences of the United States of America. 2003;100:6115-20.

[35] Rocco MA, Waraho-Zhmayev D, DeLisa MP. Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins. Proceedings of the National Academy of Sciences of the United States of America. 2012;109:13392-7.

[36] Richter S, Bruser T. Targeting of unfolded PhoA to the TAT translocon of Escherichia coli. The Journal of biological chemistry. 2005;280:42723-30.

[37] Richter S, Lindenstrauss U, Lucke C, Bayliss R, Bruser T. Functional Tat transport of unstructured, small, hydrophilic proteins. The Journal of biological chemistry. 2007;282:33257-64.

[38] Lindenstrauss U, Matos CF, Graubner W, Robinson C, Bruser T. Malfolded recombinant Tat substrates are Tat-independently degraded in Escherichia coli. FEBS letters. 2010;584:3644-8.

[39] Flanagan JM, Bewley MC. Protein quality control in bacterial cells: integrated networks of chaperones and ATP-dependent proteases. Genetic engineering. 2002;24:17-47.

[40] Meusser B, Hirsch C, Jarosch E, Sommer T. ERAD: the long road to destruction. Nature cell biology. 2005;7:766-72.

[41] Conesa A, Punt PJ, van Luijk N, van den Hondel CA. The secretion pathway in filamentous fungi: a biotechnological view. Fungal genetics and biology : FG & B. 2001;33:155-71.

[42] Fisher AC, DeLisa MP. A little help from my friends: quality control of presecretory proteins in bacteria. Journal of bacteriology. 2004;186:7467-73.

[43] Ostermeier M. Beyond cataloging the Library of Babel. Chemistry & biology. 2007;14:237-8.

[44] Tracewell CA, Arnold FH. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Current opinion in chemical biology. 2009;13:3-9.

[45] Bradbury AR, Sidhu S, Dubel S, McCafferty J. Beyond natural antibodies: the power of in vitro display technologies. Nature biotechnology. 2011;29:245-54.

[46] Waldo GS, Standish BM, Berendzen J, Terwilliger TC. Rapid protein-folding assay using green fluorescent protein. Nature biotechnology. 1999;17:691-5.

[47] Liu JW, Boucher Y, Stokes HW, Ollis DL. Improving protein solubility: the use of the Escherichia coli dihydrofolate reductase gene as a fusion reporter. Protein expression and purification. 2006;47:258-63.

[48] Maxwell KL, Mittermaier AK, Forman-Kay JD, Davidson AR. A simple in vivo assay for increased protein solubility. Protein science : a publication of the Protein Society. 1999;8:1908-11.

[49] Mansell TJ, Linderman SW, Fisher AC, DeLisa MP. A rapid protein folding assay for the bacterial periplasm. Protein science : a publication of the Protein Society. 2010;19:1079-90.

[50] Philibert P, Martineau P. 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. Microbial cell factories. 2004;3:16.

[51] Tsumoto K, Umetsu M, Kumagai I, Ejima D, Arakawa T. Solubilization of active green fluorescent protein from insoluble particles by guanidine and arginine. Biochemical and biophysical research communications. 2003;312:1383-6.

[52] Cabantous S, Terwilliger TC, Waldo GS. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nature biotechnology. 2005;23:102-7.

[53] Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht AA, Titchmarsh J, et al. Optimizing protein stability in vivo. Molecular cell. 2009;36:861-71.

[54] Wigley WC, Stidham RD, Smith NM, Hunt JF, Thomas PJ. Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. Nature biotechnology. 2001;19:131-6.

[55] Fisher AC, Kim W, DeLisa MP. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. Protein Science. 2006;15:449-58.

[56] Fisher AC, DeLisa MP. Efficient isolation of soluble intracellular single-chain antibodies using the twin-arginine translocation machinery. Journal of molecular biology. 2009;385:299-311.

[57] Fisher AC, Rocco MA, DeLisa MP. Genetic selection of solubility-enhanced proteins using the twin-arginine translocation system. Methods in molecular biology. 2011;705:53-67.

[58] Lutz S, Fast W, Benkovic SJ. A universal, vector-based system for nucleic acid reading-frame selection. Protein engineering. 2002;15:1025-30.

[59] Kim DS, Song HN, Nam HJ, Kim SG, Park YS, Park JC, et al. Directed Evolution of Human Heavy Chain Variable Domain (VH) Using In Vivo Protein Fitness Filter. PloS one. 2014;9:e98178.

[60] Wilson DB. Cellulases and biofuels. Current opinion in biotechnology. 2009;20:295-9.

[61] Wilson DB. Microbial diversity of cellulose hydrolysis. Current opinion in microbiology. 2011;14:259-63.

[62] Merino ST, Cherry J. Progress and challenges in enzyme development for biomass utilization. Advances in biochemical engineering/biotechnology. 2007;108:95-120.

[63] Wang M, Li Z, Fang X, Wang L, Qu Y. Cellulolytic enzyme production and enzymatic hydrolysis for second-generation bioethanol production. Advances in biochemical engineering/biotechnology. 2012;128:1-24.

[64] Watson DL, Wilson DB, Walker LP. Synergism in binary mixtures of Thermobifida fusca cellulases Cel6B, Cel9A, and Cel5A on BMCC and Avicel. Appl Biochem Biotechnol. 2002;101:97-111.

[65] Wilson DB. Processive and nonprocessive cellulases for biofuel production--lessons from bacterial genomes and structural analysis. Applied microbiology and biotechnology. 2012;93:497-502.

[66] Peterson R, Nevalainen H. Trichoderma reesei RUT-C30--thirty years of strain improvement. Microbiology. 2012;158:58-68.

[67] Rosgaard L, Pedersen S, Cherry JR, Harris P, Meyer AS. Efficiency of new fungal cellulase systems in boosting enzymatic degradation of barley straw lignocellulose. Biotechnology progress. 2006;22:493-8.

[68] Masai E, Katayama Y, Fukuda M. Genetic and biochemical investigations on bacterial catabolic pathways for lignin-derived aromatic compounds. Bioscience, biotechnology, and biochemistry. 2007;71:1-15.

[69] Mazzoli R, Lamberti C, Pessione E. Engineering new metabolic capabilities in bacteria: lessons from recombinant cellulolytic strategies. Trends in biotechnology. 2012;30:111-9.

[70] Percival Zhang YH, Himmel ME, Mielenz JR. Outlook for cellulase improvement: screening and selection strategies. Biotechnology advances. 2006;24:452-81.

[71] King BC, Waxman KD, Nenni NV, Walker LP, Bergstrom GC, Gibson DM. Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. Biotechnology for biofuels. 2011;4:4.

[72] Haitjema CH, Solomon KV, Henske JK, Theodorou MK, O'Malley MA. Anaerobic gut fungi: Advances in isolation, culture, and cellulolytic enzyme discovery for biofuel production. Biotechnology and bioengineering. 2014;111:1471-82.

[73] Nevalainen H, Peterson R. Making recombinant proteins in filamentous fungi- are we expecting too much? Frontiers in microbiology. 2014;5:75.

[74] Garvey M, Klose H, Fischer R, Lambertz C, Commandeur U. Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. Trends in biotechnology. 2013;31:581-93.

[75] Clarke ND. Protein engineering for bioenergy and biomass-based chemicals. Current opinion in structural biology. 2010;20:527-32.

[76] Rentmeister A. Current Developments in Cellulase Engineering. Chem-Ing-Tech. 2013;85:818-25.

[77] Bommarius AS, Sohn M, Kang Y, Lee JH, Realff MJ. Protein engineering of cellulases. Current opinion in biotechnology. 2014;29C:139-45.

[78] Guglielmi L, Denis V, Vezzio-Vie N, Bec N, Dariavach P, Larroque C, et al. Selection for intrabody solubility in mammalian cells using GFP fusions. Protein engineering, design & selection : PEDS. 2011;24:873-81.

[79] Sarkar CA, Dodevski I, Kenig M, Dudli S, Mohr A, Hermans E, et al. Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. Proceedings of the National Academy of Sciences of the United States of America. 2008;105:14808-13.

[80] Wyman CE. Twenty years of trials, tribulations, and research progress in bioethanol technology-selected key events along the way. Applied Biochemistry and Biotechnology. 2001;91-93:5-21.

[81] Rosgaard L, Pederson S, Cherry JR, Harris P, Meyer AS. Efficiency of new fungal cellulase systems in boosting enzymatic degradation of barley straw lignocellulose. Biotechnology Progress. 2006;22:493-8.

[82] Kabel MA, van der Maarel MJEC, Klip G, Voragen AGJ, Schols HA. Standard assays do not predict the efficiency of commercial cellulase preparations towards plant materials. Biotechnology and Bioengineering. 2006;93:56-63.

[83] Chundawat SP, Beckham GT, Himmel ME, Dale BE. Deconstruction of lignocellulosic biomass to fuels and chemicals. Annual review of chemical and biomolecular engineering. 2011;2:121-45.

[84] Annis SL, Goodwin PH. Recent advances in the molecular genetics of plant cell wall-degrading enzymes produced by plant pathogenic fungi. European Journal of Plant Pathology. 1997;103:1-14.

[85] Walton JD. Deconstructing the plant cell wall. Plant Physiology. 1994;104:1113-8.

[86] Tonukari NJ. Enzymes and fungal virulence. Journal of Applied Sciences & Environmental Management. 2003;7:5-8.

[87] Mendgen K, Hahn M, Deising H. Morphogenesis and mechanisms of penetration by plant pathogenic fungi. Annual Review of Phytopathology. 1996;34:367-86.

[88] Cuomo CA, Guldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science. 2007;317:1400-2.

[89] Phalip V, Delalande F, Carapito C, Goubet F, Hatsch D, Leize-Wagner E, et al. Diversity of the exoproteome of *Fusarium graminearum* grown on plant cell wall. Current Genetics. 2005;48:366-79.

[90] Paper JM, Scott-Craig JS, Adhikari ND, Cuomo CA, Walton JD. Comparative proteomics of extracellular proteins *in vitro* and *in planta* from the pathogenic fungus *Fusarium graminearum*. Proteomics. 2007;7:3171-83.

[91] Beadle BM, Shoichet BK. Structural bases of stability-function tradeoffs in enzymes. Journal of molecular biology. 2002;321:285-96.

[92] Meiering EM, Serrano L, Fersht AR. Effect of active site residues in barnase on activity and stability. Journal of molecular biology. 1992;225:585-9.

[93] Wang X, Minasov G, Shoichet BK. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. Journal of molecular biology. 2002;320:85-95.

[94] Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. Proceedings of the National Academy of Sciences of the United States of America. 2006;103:5869-74.

[95] Meinke A, Gilkes NR, Kilburn DG, Miller RC, Jr., Warren RA. Multiple domains in endoglucanase B (CenB) from Cellulomonas fimi: functions and relatedness to domains in other polypeptides. Journal of bacteriology. 1991;173:7126-35.

[96] Zhang XZ, Zhang Z, Zhu Z, Sathitsuksanoh N, Yang Y, Zhang YH. The noncellulosomal family 48 cellobiohydrolase from Clostridium phytofermentans ISDg: heterologous expression, characterization, and processivity. Applied microbiology and biotechnology. 2010;86:525-33.

[97] Mingardon F, Chanal A, Lopez-Contreras AM, Dray C, Bayer EA, Fierobe HP. Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. Applied and environmental microbiology. 2007;73:3822-32.

[98] Song JM, An YJ, Kang MH, Lee YH, Cha SS. Cultivation at 6-10 degrees C is an effective strategy to overcome the insolubility of recombinant proteins in Escherichia coli. Protein expression and purification. 2012;82:297-301.

[99] Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic acids research. 2014;42:D490-5.

[100] Hong J, Tamaki H, Akiba S, Yamamoto K, Kumagai H. Cloning of a gene encoding a highly stable endo-beta-1,4-glucanase from *Aspergillus niger* and its expression in yeast. Journal of Bioscience and Bioengineering. 2001;92:434-41.

[101] King BC, Donnelly MK, Bergstrom GC, Walker LP, Gibson DM. An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. Biotechnol Bioeng. 2009;102:1033-44.

[102] Ghose TK. Measurement of Cellulase Activities. Pure and Applied Chemistry. 1987;59:257-68.

[103] Tokuriki N, Stricher F, Serrano L, Tawfik DS. How protein stability and new functions trade off. PLoS computational biology. 2008;4:e1000002.

[104] Walton J, Banerjee G, Car S. GENPLAT: an automated platform for biomass enzyme discovery and cocktail optimization. Journal of visualized experiments : JoVE. 2011.

[105] Foo JL, Ching CB, Chang MW, Leong SS. The imminent role of protein engineering in synthetic biology. Biotechnology advances. 2012;30:541-9.

[106] Yoshikuni Y, Dietrich JA, Nowroozi FF, Babbitt PC, Keasling JD. Redesigning enzymes based on adaptive evolution for optimal function in synthetic metabolic pathways. Chemistry & biology. 2008;15:607-18.

[107] Fromant M, Blanquet S, Plateau P. Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. Analytical biochemistry. 1995;224:347-53.

[108] Farrow MF, Arnold FH. High throughput screening of fungal endoglucanase activity in Escherichia coli. Journal of visualized experiments : JoVE. 2011.

[109] Monsellier E, Bedouelle H. Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength. Protein engineering, design & selection : PEDS. 2005;18:445-56.

[110] Pace CN. Measuring and increasing protein stability. Trends in biotechnology. 1990;8:93-8.

[111] Robinson C, Matos CF, Beck D, Ren C, Lawrence J, Vasisht N, et al. Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. Biochimica et biophysica acta. 2011;1808:876-84.

[112] Kostecki JS, Li H, Turner RJ, DeLisa MP. Visualizing interactions along the Escherichia coli twin-arginine translocation pathway using protein fragment complementation. PloS one. 2010;5:e9225.

[113] Rocco MA. Twin-arginine translocase Mutations that supress folding quality control and permit export of misfolded substrate proteins. Dissertation. 2012.

[114] De Leeuw E, Porcelli I, Sargent F, Palmer T, Berks BC. Membrane interactions and self-association of the TatA and TatB components of the twin-arginine translocation pathway. FEBS letters. 2001;506:143-8.

[115] Maldonado B, Kneuper H, Buchanan G, Hatzixanthis K, Sargent F, Berks BC, et al. Characterisation of the membrane-extrinsic domain of the TatB component of the twin arginine protein translocase. FEBS letters. 2011;585:478-84.

[116] Lee PA, Buchanan G, Stanley NR, Berks BC, Palmer T. Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation. Journal of bacteriology. 2002;184:5871-9.

[117] Kneuper H, Maldonado B, Jager F, Krehenbrink M, Buchanan G, Keller R, et al. Molecular dissection of TatC defines critical regions essential for protein transport and a TatB-TatC contact site. Molecular microbiology. 2012;85:945-61.

[118] Zoufaly S, Frobel J, Rose P, Flecken T, Maurer C, Moser M, et al. Mapping precursor-binding site on TatC subunit of twin arginine-specific protein translocase by site-specific photo cross-linking. The Journal of biological chemistry. 2012;287:13430-41.

[119] Lee PA, Orriss GL, Buchanan G, Greene NP, Bond PJ, Punginelli C, et al. Cysteine-scanning mutagenesis and disulfide mapping studies of the conserved domain of the twin-arginine translocase TatB component. The Journal of biological chemistry. 2006;281:34072-85.

[120] Buchner J, Grallert H, Jakob U. Analysis of chaperone function using citrate synthase as nonnative substrate protein. Methods in enzymology. 1998;290:323-38.

[121] Wang JD, Herman C, Tipton KA, Gross CA, Weissman JS. Directed evolution of substrate-optimized GroEL/S chaperonins. Cell. 2002;111:1027-39.

[122] Missiakas D, Georgopoulos C, Raina S. The Escherichia coli dsbC (xprA) gene encodes a periplasmic protein involved in disulfide bond formation. The EMBO journal. 1994;13:2013-20.

[123] Matos CF, Di Cola A, Robinson C. TatD is a central component of a Tat translocon-initiated quality control system for exported FeS proteins in Escherichia coli. EMBO reports. 2009;10:474-9.

[124] Eser M, Ehrmann M. SecA-dependent quality control of intracellular protein localization. Proceedings of the National Academy of Sciences of the United States of America. 2003;100:13231-4.

[125] Fisher AC, Kim W, DeLisa MP. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. Protein science : a publication of the Protein Society. 2006;15:449-58.

[126] Waraho D, DeLisa MP. Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. Proceedings of the National Academy of Sciences of the United States of America. 2009;106:3692-7.

[127] Karlsson AJ, Lim HK, Xu H, Rocco MA, Bratkowski MA, Ke A, et al. Engineering antibody fitness and function using membrane-anchored display of correctly folded proteins. Journal of molecular biology. 2012;416:94-107.

[128] Bruser T. The twin-arginine translocation system and its capability for protein secretion in biotechnological protein production. Applied microbiology and biotechnology. 2007;76:35-45.

[129] Stanley NR, Sargent F, Buchanan G, Shi J, Stewart V, Palmer T, et al. Behaviour of topological marker proteins targeted to the Tat protein transport pathway. Molecular microbiology. 2002;43:1005-21.

[130] Hicks MG, Lee PA, Georgiou G, Berks BC, Palmer T. Positive selection for loss-of-function tat mutations identifies critical residues required for TatA activity. Journal of bacteriology. 2005;187:2920-5.

[131] Blaudeck N, Kreutzenbeck P, Muller M, Sprenger GA, Freudl R. Isolation and characterization of bifunctional Escherichia coli TatA mutant proteins that allow efficient

tat-dependent protein translocation in the absence of TatB. The Journal of biological chemistry. 2005;280:3426-32.

[132] Barrett CM, Freudl R, Robinson C. Twin arginine translocation (Tat)-dependent export in the apparent absence of TatABC or TatA complexes using modified Escherichia coli TatA subunits that substitute for TatB. The Journal of biological chemistry. 2007;282:36206-13.

[133] Mizuno T. Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of Escherichia coli. DNA research : an international journal for rapid publication of reports on genes and genomes. 1997;4:161-8.

[134] Wexler M, Sargent F, Jack RL, Stanley NR, Bogsch EG, Robinson C, et al. TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in sec-independent protein export. The Journal of biological chemistry. 2000;275:16717-22.

[135] Harris PV, Welner D, McFarland KC, Re E, Navarro Poulsen JC, Brown K, et al. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. Biochemistry. 2010;49:3305-16.

[136] Karkehabadi S, Hansson H, Kim S, Piens K, Mitchinson C, Sandgren M. The first structure of a glycoside hydrolase family 61 member, Cel61B from Hypocrea jecorina, at 1.6 A resolution. Journal of molecular biology. 2008;383:144-54.

[137] Gusakov AV. Alternatives to Trichoderma reesei in biofuel production. Trends in biotechnology. 2011;29:419-25.

[138] Kautto L, Grinyer J, Paulsen I, Tetu S, Pillai A, Pardiwalla S, et al. Stress effects caused by the expression of a mutant cellobiohydrolase I and proteasome inhibition in Trichoderma reesei Rut-C30. New biotechnology. 2013;30:183-91.

[139] Beckham GT, Dai Z, Matthews JF, Momany M, Payne CM, Adney WS, et al. Harnessing glycosylation to improve cellulase activity. Current opinion in biotechnology. 2012;23:338-45.

[140] Bessette PH, Aslund F, Beckwith J, Georgiou G. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. Proceedings of the National Academy of Sciences of the United States of America. 1999;96:13703-8.

[141] Druzhinina IS, Shelest E, Kubicek CP. Novel traits of Trichoderma predicted through the analysis of its secretome. FEMS microbiology letters. 2012;337:1-9.

[142] Svergun DI, Koch MHJ. Small-angle scattering studies of biological macromolecules in solution. Rep Prog Phys. 2003;66:1735-82.

[143] Receveur V, Czjzek M, Schulein M, Panine P, Henrissat B. Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. The Journal of biological chemistry. 2002;277:40887-92.

[144] Bianchetti CM, Brumm P, Smith RW, Dyer K, Hura GL, Rutkoski TJ, et al. Structure, dynamics, and specificity of endoglucanase D from Clostridium cellulovorans. Journal of molecular biology. 2013;425:4267-85.

[145] Lima MA, Oliveira-Neto M, Kadowaki MA, Rosseto FR, Prates ET, Squina FM, et al. Aspergillus niger beta-glucosidase has a cellulase-like tadpole molecular shape: insights into glycoside hydrolase family 3 (GH3) beta-glucosidase structure and function. The Journal of biological chemistry. 2013;288:32991-3005.

[146] Semenyuk AV, Svergun DI. Gnom - a Program Package for Small-Angle Scattering Data-Processing. Journal of applied crystallography. 1991;24:537-40.

[147] Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of applied crystallography. 2009;42:342-6.

[148] Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. Journal of applied crystallography. 2003;36:860-4.

[149] Stals I, Sandra K, Geysens S, Contreras R, Van Beeumen J, Claeyssens M. Factors influencing glycosylation of Trichoderma reesei cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. Glycobiology. 2004;14:713-24.

[150] Khalikova E, Susi P, Korpela T. Microbial dextran-hydrolyzing enzymes: fundamentals and applications. Microbiology and molecular biology reviews : MMBR. 2005;69:306-25.

[151] Shimizu E, Unno T, Ohba M, Okada G. Purification and characterization of an isomaltotriose-producing endo-dextranase from a Fusarium sp. Bioscience, biotechnology, and biochemistry. 1998;62:117-22.

[152] Nielsen SS, Moller M, Gillilan RE. High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell. Journal of applied crystallography. 2012;45:213-23.

[153] Kufer P, Lutterbuse R, Baeuerle PA. A revival of bispecific antibodies. Trends Biotechnol. 2004;22:238-44.

[154] Gruber M, Schodin BA, Wilson ER, Kranz DM. Efficient tumor cell lysis mediated by a bispecific single chain antibody expressed in Escherichia coli. Journal of immunology. 1994;152:5368-74.

[155] Mallender WD, Voss EW, Jr. Construction, expression, and activity of a bivalent bispecific single-chain antibody. J Biol Chem. 1994;269:199-206.

[156] Heck T, Faccio G, Richter M, Thony-Meyer L. Enzyme-catalyzed protein crosslinking. Applied microbiology and biotechnology. 2013;97:461-75.

[157] Tanaka T, Kamiya N, Nagamune T. Peptidyl linkers for protein heterodimerization catalyzed by microbial transglutaminase. Bioconjugate chemistry. 2004;15:491-7.

[158] Schneewind O, Model P, Fischetti VA. Sorting of protein A to the staphylococcal cell wall. Cell. 1992;70:267-81.

[159] Popp MW, Ploegh HL. Making and breaking peptide bonds: protein engineering using sortase. Angewandte Chemie. 2011;50:5024-32.

[160] Proft T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. Biotechnology letters. 2010;32:1-10.

[161] Ritzefeld M. Sortagging: a robust and efficient chemoenzymatic ligation strategy. Chemistry. 2014;20:8516-29.

[162] Tsukiji S, Nagamune T. Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering. Chembiochem : a European journal of chemical biology. 2009;10:787-98.

[163] Parthasarathy R, Subramanian S, Boder ET. Sortase A as a novel molecular "stapler" for sequence-specific protein conjugation. Bioconjugate chemistry. 2007;18:469-76.

[164] Clow F, Fraser JD, Proft T. Immobilization of proteins to biacore sensor chips using Staphylococcus aureus sortase A. Biotechnology letters. 2008;30:1603-7.

[165] Tanaka T, Yamamoto T, Tsukiji S, Nagamune T. Site-specific protein modification on living cells catalyzed by Sortase. Chembiochem : a European journal of chemical biology. 2008;9:802-7.

[166] Wu S, Proft T. The use of sortase-mediated ligation for the immobilisation of bacterial adhesins onto fluorescence-labelled microspheres: a novel approach to analyse bacterial adhesion to host cells. Biotechnology letters. 2010;32:1713-8.

[167] Antos JM, Popp MW, Ernst R, Chew GL, Spooner E, Ploegh HL. A straight path to circular proteins. The Journal of biological chemistry. 2009;284:16028-36.

[168] Popp MW, Dougan SK, Chuang TY, Spooner E, Ploegh HL. Sortase-catalyzed transformations that improve the properties of cytokines. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:3169-74.

[169] Strijbis K, Ploegh HL. Secretion of circular proteins using sortase. Methods in molecular biology. 2014;1174:73-83.

[170] Strijbis K, Spooner E, Ploegh HL. Protein ligation in living cells using sortase. Traffic. 2012;13:780-9.

[171] Zhang J, Yamaguchi S, Hirakawa H, Nagamune T. Intracellular protein cyclization catalyzed by exogenously transduced Streptococcus pyogenes sortase A. Journal of bioscience and bioengineering. 2013;116:298-301.

[172] Bentley ML, Gaweska H, Kielec JM, McCafferty DG. Engineering the substrate specificity of Staphylococcus aureus Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. The Journal of biological chemistry. 2007;282:6571-81.

[173] Piotukh K, Geltinger B, Heinrich N, Gerth F, Beyermann M, Freund C, et al. Directed evolution of sortase A mutants with altered substrate selectivity profiles. Journal of the American Chemical Society. 2011;133:17536-9.

[174] Levary DA, Parthasarathy R, Boder ET, Ackerman ME. Protein-protein fusion catalyzed by sortase A. PloS one. 2011;6:e18342.

[175] Huang X, Aulabaugh A, Ding W, Kapoor B, Alksne L, Tabei K, et al. Kinetic mechanism of Staphylococcus aureus sortase SrtA. Biochemistry. 2003;42:11307-15.

[176] Bentley ML, Lamb EC, McCafferty DG. Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from Staphylococcus aureus. The Journal of biological chemistry. 2008;283:14762-71.

[177] Kruger RG, Otvos B, Frankel BA, Bentley M, Dostal P, McCafferty DG. Analysis of the substrate specificity of the Staphylococcus aureus sortase transpeptidase SrtA. Biochemistry. 2004;43:1541-51.

[178] Naik MT, Suree N, Ilangovan U, Liew CK, Thieu W, Campbell DO, et al. Staphylococcus aureus Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. The Journal of biological chemistry. 2006;281:1817-26.

[179] Hirakawa H, Ishikawa S, Nagamune T. Design of Ca2+-independent Staphylococcus aureus sortase A mutants. Biotechnology and bioengineering. 2012;109:2955-61.

[180] Suree N, Liew CK, Villareal VA, Thieu W, Fadeev EA, Clemens JJ, et al. The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. The Journal of biological chemistry. 2009;284:24465-77.

[181] Chen I, Dorr BM, Liu DR. A general strategy for the evolution of bond-forming enzymes using yeast display. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:11399-404.

[182] Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M. A genetic system yields self-cleaving inteins for bioseparations. Nature biotechnology. 1999;17:889-92.

[183] Wu WY, Mee C, Califano F, Banki R, Wood DW. Recombinant protein purification by self-cleaving aggregation tag. Nature protocols. 2006;1:2257-62.

[184] Nguyen AW, Daugherty PS. Evolutionary optimization of fluorescent proteins for intracellular FRET. Nature biotechnology. 2005;23:355-60.

[185] Banki MR, Wood DW. Inteins and affinity resin substitutes for protein purification and scale up. Microbial cell factories. 2005;4:32.

[186] You X, Nguyen AW, Jabaiah A, Sheff MA, Thorn KS, Daugherty PS. Intracellular protein interaction mapping with FRET hybrids. Proceedings of the National Academy of Sciences of the United States of America. 2006;103:18458-63.

[187] Koch H, Grafe N, Schiess R, Pluckthun A. Direct selection of antibodies from complex libraries with the protein fragment complementation assay. Journal of molecular biology. 2006;357:427-41.

[188] Martineau P, Jones P, Winter G. Expression of an antibody fragment at high levels in the bacterial cytoplasm. Journal of molecular biology. 1998;280:117-27.

[189] Dalboge H, Bayne S, Pedersen J. In vivo processing of N-terminal methionine in E. coli. FEBS letters. 1990;266:1-3.

[190] Varshavsky A. The N-end rule pathway of protein degradation. Genes to cells : devoted to molecular & cellular mechanisms. 1997;2:13-28.

[191] Ben-Bassat A, Bauer K, Chang SY, Myambo K, Boosman A, Chang S. Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure. Journal of bacteriology. 1987;169:751-7.

[192] Gangola P, Rosen BP. Maintenance of intracellular calcium in Escherichia coli. The Journal of biological chemistry. 1987;262:12570-4.

[193] Dzivenu OK, Park HH, Wu H. General co-expression vectors for the overexpression of heterodimeric protein complexes in Escherichia coli. Protein expression and purification. 2004;38:1-8.

[194] Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Nature reviews Molecular cell biology. 2009;10:866-76.

[195] Bailey JE. Toward a science of metabolic engineering. Science. 1991;252:1668-75.

[196] Stephanopoulos G, Vallino JJ. Network rigidity and metabolic engineering in metabolite overproduction. Science. 1991;252:1675-81.

[197] Nielsen J, Fussenegger M, Keasling J, Lee SY, Liao JC, Prather K, et al. Engineering synergy in biotechnology. Nature chemical biology. 2014;10:319-22.

[198] Prather KL, Martin CH. De novo biosynthetic pathways: rational design of microbial chemical factories. Current opinion in biotechnology. 2008;19:468-74.

[199] Atsumi S, Liao JC. Metabolic engineering for advanced biofuels production from Escherichia coli. Current opinion in biotechnology. 2008;19:414-9.

[200] Mukhopadhyay A, Redding AM, Rutherford BJ, Keasling JD. Importance of systems biology in engineering microbes for biofuel production. Current opinion in biotechnology. 2008;19:228-34.

[201] Zhang F, Rodriguez S, Keasling JD. Metabolic engineering of microbial pathways for advanced biofuels production. Current opinion in biotechnology. 2011;22:775-83.

[202] Keasling JD. Synthetic biology and the development of tools for metabolic engineering. Metabolic engineering. 2012;14:189-95.

[203] Kung Y, Runguphan W, Keasling JD. From fields to fuels: recent advances in the microbial production of biofuels. ACS synthetic biology. 2012;1:498-513.

[204] Chemler JA, Koffas MA. Metabolic engineering for plant natural product biosynthesis in microbes. Current opinion in biotechnology. 2008;19:597-605.

[205] Winter JM, Tang Y. Synthetic biological approaches to natural product biosynthesis. Current opinion in biotechnology. 2012;23:736-43.

[206] Pickens LB, Tang Y, Chooi YH. Metabolic engineering for the production of natural products. Annual review of chemical and biomolecular engineering. 2011;2:211-36.

[207] Nakamura CE, Whited GM. Metabolic engineering for the microbial production of 1,3-propanediol. Current opinion in biotechnology. 2003;14:454-9.

[208] Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature. 2008;451:86-9.

[209] Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB. Microbial biosynthesis of alkanes. Science. 2010;329:559-62.

[210] Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature. 2006;440:940-3.

[211] Way JC, Collins JJ, Keasling JD, Silver PA. Integrating biological redesign: where synthetic biology came from and where it needs to go. Cell. 2014;157:151-61.

[212] Zabala AO, Cacho RA, Tang Y. Protein engineering towards natural product synthesis and diversification. Journal of industrial microbiology & biotechnology. 2012;39:227-41.

[213] Yadav VG, De Mey M, Lim CG, Ajikumar PK, Stephanopoulos G. The future of metabolic engineering and synthetic biology: towards a systematic practice. Metabolic engineering. 2012;14:233-41.

[214] Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nature biotechnology. 2006;24:1027-32.

[215] Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nature biotechnology. 2009;27:946-50.

[216] Tyo KE, Ajikumar PK, Stephanopoulos G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nature biotechnology. 2009;27:760-5.

[217] Alper H, Miyaoku K, Stephanopoulos G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nature biotechnology. 2005;23:612-6.

[218] Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, Tidor B, et al. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. Proceedings of the National Academy of Sciences of the United States of America. 2010;107:13654-9.

[219] Zhang K, Nelson KM, Bhuripanyo K, Grimes KD, Zhao B, Aldrich CC, et al. Engineering the substrate specificity of the DhbE adenylation domain by yeast cell surface display. Chemistry & biology. 2013;20:92-101.

[220] Fasan R, Meharenna YT, Snow CD, Poulos TL, Arnold FH. Evolutionary history of a specialized p450 propane monooxygenase. Journal of molecular biology. 2008;383:1069-80.

[221] Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, et al. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America. 2013;110:14024-9.

[222] Keasling J. Advanced Fuels from Advanced Plants. Cornell CBI Seminar. 2012.

[223] Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY, Lebar T, et al. DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. Nucleic acids research. 2012;40:1879-89.

[224] Baltz RH. Molecular engineering approaches to peptide, polyketide and other antibiotics. Nature biotechnology. 2006;24:1533-40.

[225] Khosla C. Structures and mechanisms of polyketide synthases. The Journal of organic chemistry. 2009;74:6416-20.

[226] Poust S, Hagen A, Katz L, Keasling JD. Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. Current opinion in biotechnology. 2014;30C:32-9.

[227] Pfeifer BA, Khosla C. Biosynthesis of polyketides in heterologous hosts. Microbiology and molecular biology reviews : MMBR. 2001;65:106-18.

[228] Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, et al. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nature biotechnology. 2005;23:1171-6.

[229] Menzella HG, Carney JR, Li Y, Santi DV. Using chemobiosynthesis and synthetic mini-polyketide synthases to produce pharmaceutical intermediates in Escherichia coli. Applied and environmental microbiology. 2010;76:5221-7.

[230] Gaj T, Gersbach CA, Barbas CF, 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in biotechnology. 2013;31:397-405.

[231] Lynch SA, Gill RT. Synthetic biology: new strategies for directing design. Metabolic engineering. 2012;14:205-11.

[232] Johnson ET, Baron DB, Naranjo B, Bond DR, Schmidt-Dannert C, Gralnick JA. Enhancement of survival and electricity production in an engineered bacterium by light-driven proton pumping. Applied and environmental microbiology. 2010;76:4123-9.

[233] Berla BM, Saha R, Immethun CM, Maranas CD, Moon TS, Pakrasi HB. Synthetic biology of cyanobacteria: unique challenges and opportunities. Frontiers in microbiology. 2013;4:246.

[234] Atsumi S, Higashide W, Liao JC. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nature biotechnology. 2009;27:1177-80.