In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing material of this type. I agree that permission to copy from, or to publish, this dissertation may be granted by the Professor under whose direction it was written, or in his absence, by the Dean of the Graduate School when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

_____

Holly E. Carpenter

Biomaterials Design: Creation of a novel orthogonal translational system in *E. coli* for the site-specific incorporation of proline analogues in biosynthetic protein materials

By

Holly E. Carpenter
Doctor of Philosophy

Department of Chemistry

_____
Vincent P. Conticello, Ph.D.
Adviser

_____
David G. Lynn, Ph.D.
Committee Member

_____
Dale E. Edmondson, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

_____
Date

Biomaterials Design: Creation of a novel orthogonal translational system in *E. coli* for the site-specific incorporation of proline analogues in biosynthetic protein materials

By

Holly E. Carpenter
B.S., North Georgia College & State University, 2002

Adviser

Vincent P. Conticello, Ph.D.

An Abstract of
A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

2008

Abstract

New techniques have recently been presented that enable co-translational, *in vivo* incorporation of amino acid analogues in proteins through an expansion of the genetic code. In general terms, methods to expand the genetic code have included the selection and directed evolution of *orthogonal* tRNA/aminoacyl-tRNA synthetase (aaRS) pairs that function in a heterologous host, such that a non-canonical amino acid is inserted in response to a termination codon using non-sense suppression. While successful incorporation of non-canonical amino acids has been achieved, non-sense suppression has limited efficiency, particularly for multi-site incorporation of amino acid analogues. Therefore, a different approach was chosen that involved capture of codons that are rarely utilized in the native host organism. Given this framework, the CCC codon of the proline family box was chosen as a target for incorporation of non-canonical proline analogues in an *E. coli* expression system where proline or proline analogues were directly incorporated at the site of multiple CCC codons to produce an elastin-mimetic protein material. Chapter 1 gives a broad introduction to relevant topic areas, while Chapter 2 details the specific design and genetic engineering of a gene for use in production of the elastin-mimetic protein. Production of an elastin-mimetic material serves as a model for synthetic collagen biosynthesis where proline and hydroxyproline are directly incorporated at multiple sites into the same polypeptide.

Chapter 3 describes the design, genetic engineering, and evaluation of a novel prolyl-tRNA synthetase/$tRNA^{Pro}$ pair, from *M. jannaschii*, that functions as an orthogonal pair in *E. coli*. The pair was used to demonstrate orthogonality in the host and to produce

an elastin-mimetic protein with incorporation of proline and proline analogues at specific sites throughout the primary sequence of the polypeptide (~ 80 CCC sites).

Chapter 4 gives an in-depth account of the development of directed evolution techniques used for engineering the MjProRS enzyme for incorporation of (2*S*,4*R*)-4-hydroxyproline in elastin-mimetic proteins. Library construction of MjProRS variants and high-throughput screening of the libraries using a fluorescence-activated cell sorting (FACS)-based method were optimized for use in isolation of a synthetase mutant that would be specific for (2*S*,4*R*)-4-hydroxyproline. Design of a synthetase enzyme that exhibited a preference for the imino acid analogue over the natural proline substrate would be a significant advance toward the design of synthetic collagen production where proline and hydroxyproline incorporation would occur simultaneously in a bacterial host using separate aminoacyl-tRNA synthetase/tRNA pairs.

Chapter 5 outlines the design and expression of recombinant human tropoelastin in *E. coli*. Human tropoelastin is an attractive biomaterial target as it is the natural soluble precursor to elastin in higher mammals. Extensive cross-linking occurs between lysine residues in the tropoelastin sequence to form mature elastin *in vivo*. Recombinant tropoelastin production in bacteria offers a cost-effective and high volume production method for use of tropoelastin or tropoelastin variants (with amino acid analogue incorporation) in biomaterial applications.

Biomaterials Design: Creation of a novel orthogonal translational system in *E. coli* for the site-specific incorporation of proline analogues in biosynthetic protein materials

By

Holly E. Carpenter
B.S., North Georgia College & State University, 2002

Adviser

Vincent P. Conticello, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

2008

Acknowledgements

I would first like to thank my adviser, Vince Conticello, for his steady support and encouragement throughout my graduate career. I greatly appreciate his enthusiasm for science and for his interest in helping me to pursue a career in academia. The research skills that I have learned through his guidance will certainly help me throughout my career and throughout life. I have so many fond memories of our talks in the lab about science and life in general. He is truly a wonderful mentor and friend. Thank you so much Vince.

Thank you also to my committee members, Dr. Lynn and Dr. Edmondson, for their helpful guidance. I will never forget the first time I sent out my original research proposal for my committee to approve. Dr. Lynn asked me to stop by his office and he gave me great advice on the proposal topic. He took time out of his busy schedule, as the Department Chair, to help me—there are not many other programs where the faculty care so much about the education of the students as they do at Emory. Also, I remember sitting down in a meeting with Dr. Edmonson to talk about my career plans in which he listened to my goals and encouraged me with kind words of support. Even if he is out of town, he is always the first to reply to me when I contact him about letters of support and anything else for that matter!

Many, many thanks to Pat Marsteller for being a fantastic mentor and guide from the very beginning. From SURE to PRISM (I and II) to HHMI, I have taken full advantage of so many programs that Pat has built at Emory. She has been a guiding force in my career as well as a great friend. From her inspiring career to her witty comments and everything in between, Pat truly deserves a medal of honor. I cannot forget all of our

fun moments together, from meeting the President in Perú, to pool parties in the backyard. Thank you, Pat, for all that you have done for me and so many other students.

To my mentors at North Georgia, Dr. Brad Herbert and Dr. Ralph Hitt, thank you for always being there for me and taking the time to help me. I could fill a book with all that you have done for me and for so many others. Brad, I could never forget all of the fun times and jokes we have shared together at Furman and at North Georgia over the past 9 years. Without your help and encouragement, I would not have finally become "a scientist". Ralph, you are truly an inspiration to us all in your kindness, thoughtfulness, and supportive manner. I remember all of the times we have shared together over the years and how you always made me feel confident in my studies and in my career when nothing else seemed to be going right. Thank you also to my friends, Dr. Tom Fox, Dr. Mark Davis, Dr. Frank Corotto, Dr. Wheeler, and K.M. for so many fond memories!

A huge thank you to Bob Karaffa for his technical expertise and helpful insight. Bob is a joy to work with, and his FACS sorting skills are beyond compare. Over the years, I have been blessed to work with wonderful people in the lab. We are all part of the "Conticello Family". Thanks to Jaimie Anderson, Yuri Zimenkov, Wookhyun Kim, Jet Qi, Melissa Patterson and Sonha Payne for teaching me how to do most everything in the lab (especially Jet!). Thank you to all of you (my lab mates) who have made working in the lab such a wonderful experience.

Thanks to all of my wonderful friends who have listened to me, supported me, and encouraged me over these long years. To Kasey (my maid of honor), Kirklyn, Uliana, Rory, Melissa, Phyllis, Kish, Earnest, Kim, Naomi, Adrienne, Jordan, Tiffany, Carolina, Matt, Cameron, Betül, Shana, Sean, Steven, and many others, thank you all.

A special thanks to my parents, Sarah and Steve. Through it all, your love has seen me through and in the darkest times, you have always shown me the light. Your support is a blessing and I appreciate all that you both have done. You taught me to never give up and you always encouraged me in all that I set out to do. It is because of you, in such a large part, that I am here today and that I have been able to accomplish my goals. Thank you. Thank you also to my brother, Jim. What can I say? You are so special to me in so many ways and I thank you for being you. We will always share the bonds of family, friendship, love, and understanding that keep us strong. To my special Nan, I love to think back on all of our fun memories. There are too many to count from cartoons to graduations, your guidance and wisdom (and keen fashion sense) carried me through it all. Thanks also to Granny, Gran, and Tom who are no longer with us but certainly here in spirit and in memory as their strength in life was an inspiration to us all.

To the one I thank most of all, the love of my life, my fiancé, Shawn Desai. Shawn has been the foundation of my strength, my knight in shining armor, and I would follow him to the ends of the earth. He is the most kind and caring man I have ever met and I look forward to the life that we will share together. I dedicate this thesis to Shawn, my soulmate and one true love.

Farewell we call to hearth and hall!

Though wind may blow and rain may fall,

We must away ere break of day

Far over wood and mountain tall

To answer next our ensuing Call.

Modified from a verse in The Fellowship of the Ring,
by J.R.R. Tolkien

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 3

Chapter 4

Chapter 5

List of Schemes

Chapter 1

Chapter 2

Chapter 4

Chapter 5

Chapter 1

Introduction

Biomaterials: Elastin and collagen

As the basic building blocks of life, cells utilize molecular machinery to carry out complex processes that make growth, development, and survival possible. Proteins hold a position of particular importance as catalysts and invaluable structural elements of biological systems (Fig. 1)[1]. Native elastin, for example, is widely distributed in vertebrate tissues and provides elasticity and resiliency to tissues and organs such as the arteries of the cardiovascular system, the lungs, and the skin[2]. Elastin and collagen are the primary protein components of connective tissue in vertebrate organisms, and elastin's resiliency, combined with the strength of collagen, confers durability in repetitively extended arterial walls and in the highly used tissues of the lung and skin[3-5]. Mature, native elastin has a half-life of 70 years and is among the most stable proteins known[6]. Elastin and collagen are ideal targets for laboratory-based, biomaterial design and engineering applications due to the highly useful and remarkable physical properties of these natural biomaterials[7].

Evidence suggests that elastin is encoded by a single gene present in the mammalian genome[8], yet multiple tropoelastin isoforms exist in the human body due to alternative splicing of the 34 exons that comprise the gene[3,9-12]. The natural, soluble precursor of native elastin, tropoelastin, consists of alternating hydrophobic domains rich in non-polar amino acids and hydrophilic domains rich in lysine residues[3]. The hydrophobic domains, believed to be important for the elastic properties of tropoelastin, are characterized by three to six peptide repeats with sequences such as GVGVP, GGVP, and GVGVAP. The hydrophilic domains, involved in cross-linking, consist of stretches

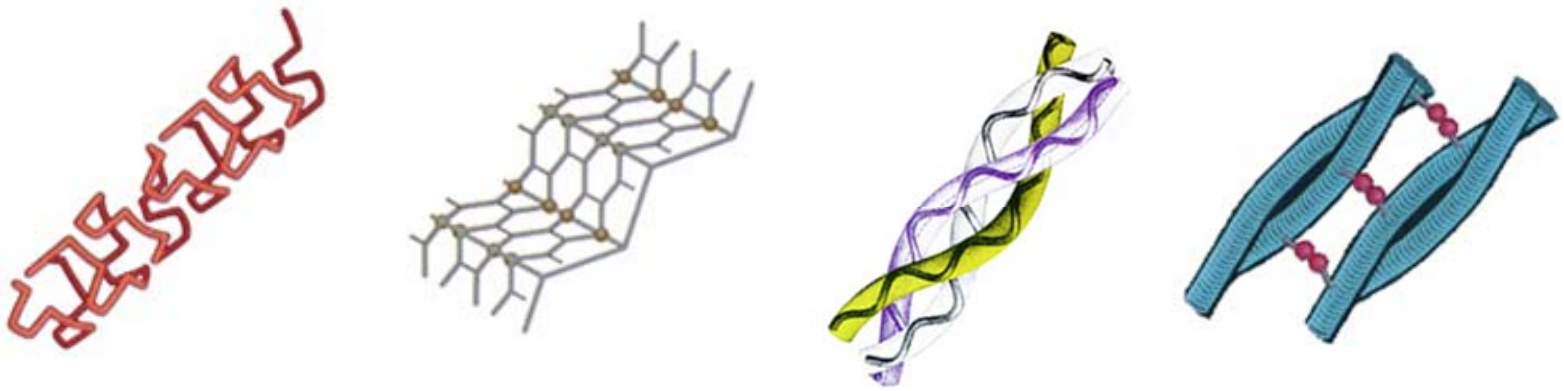


Figure 1. Natural protein polymers are not only useful in biomaterials applications, but also they perform a variety of important functions in biological systems.

of lysine residues separated by alanine residues such as in the sequence AAAKAAKAAA. In some tropoelastin isoforms, proline is present in place of a polyAla tract[13]. The hydrophilic domains display α-helical character while the hydrophobic domains exist in a random coil conformation with evidence for the presence of type II like β-turns within the structure[14]. Tropoelastin is rich in glycine (33%), proline (10-13%), and other hydrophobic residues (44%) giving elastin a unique and largely hydrophobic amino acid composition[15]. The hydrophobic domain sequences among species are evolutionarily divergent, while the hydrophilic, cross-linking domain sequences have been highly conserved over time possibly due to the conformational constraints of cross-linking[11].

In mammals, tropoelastin is synthesized with an N-terminal 26 amino acid secretion signal sequence that allows for secretion of tropoelastin from the cells as a ~72 kDa protein[3,15]. Deposition of tropoelastin in the extracellular matrix is a highly complex process that involves protein binding factors and subsequent fibrillogenesis. *In vivo*, elastin fibers appear to exist as two morphologically different components, a highly isotropic amorphous elastin constituent within an organized microfibrillar, cross-linked scaffold primarily composed of fibrillin[3]. The complex mechanism of fiber assembly in native, mature elastin is not well understood. Coacervation is one interesting property of soluble tropoelastin that is believed to be important for fibril formation[16]. Upon increasing the temperature of a solution of soluble tropoelastin above 20 ºC, the solution becomes turbid due to self-assembly and aggregation of tropoelastin molecules. Aggregation is thermodynamically controlled, and is therefore fully reversible by cooling

the solution[16]. Coacervation is believed to be caused by interaction between the hydrophobic domains, and plays an important role in fibrillogenesis[2,16-19].

Coacervation is also a desirable property of elastin-mimetic polypeptides developed for biomaterials applications as the self-assembly and subsequent insolubility of elastin-mimetics at temperatures above the intrinsic, inverse temperature transition allow for development of elastin-based hydrogels, materials, and devices useful in biomedical applications[1,7]. Although mature elastin is insoluble in the native form, elastin-mimetic polypeptides have been developed that are soluble mimics of native elastin. Consensus tetra- (VPGG), penta- (VPGVG), and hexapeptide (APGVGV) repeat motifs can be used in biomaterials to mimic the characteristic properties of native elastin[20-23]. Urry and coworkers have made extensive progress that provides a framework for further study using synthetic elastin with the pentapeptide repeat sequence, Val-Pro-Gly-Xaa-Gly[24]. The Xaa position can be occupied by any amino acid (often valine or isoleucine), and the identity of the amino acid has a significant impact on the overall physical properties of elastin-mimetic polypeptides. The fourth position, in particular, has been shown to play an important role in modulating the inverse temperature transition of elastomeric materials with more polar amino acids increasing the transition temperature (Fig. 2)[25].

Conticello and coworkers, among others, have developed genetic engineering strategies for the production of elastin-mimetic biomaterials[23,24,26-31]. Tropoelastin and elastin-mimetics are particularly useful in laboratory applications in that they display the characteristic physical properties of native elastin, but they are soluble at room

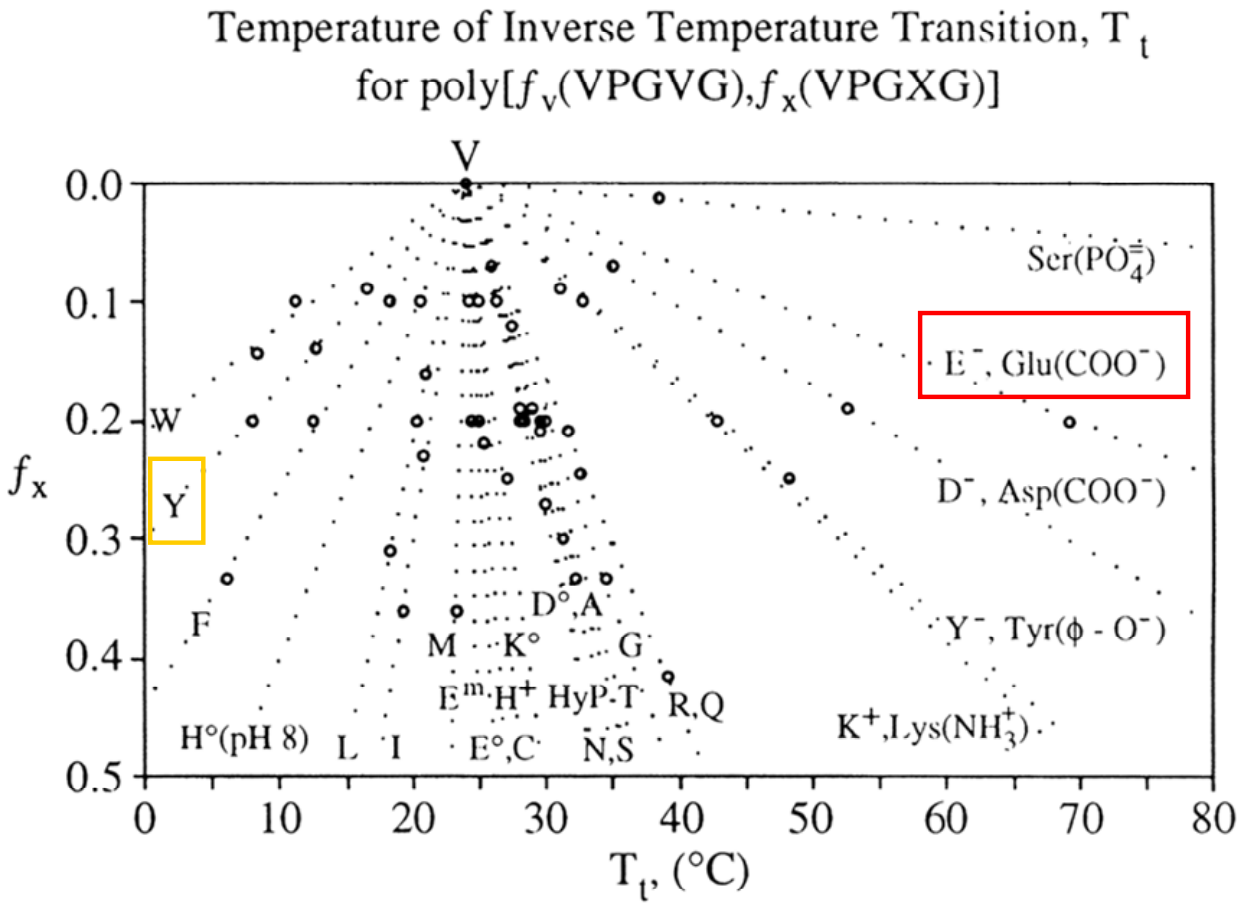


Figure 2. Characterization of elastin-mimetic polypeptides. Substitution of valine 4 with hydrophilic or hydrophobic amino acid residues causes an increase or a decrease in the transition temperature (*T*t), respectively. Hydrophobicity scale, where *f*v and *f*x are the mole fractions of each peptide and *f*v + *f*x = 1.

temperature and at temperatures below the inverse temperature transition[3]. Protein polymers, based on the pentapeptide sequence (VPGVG), exhibit elastic behavior with spectroscopic features that are consistent with those of native elastin, including a highly mobile backbone and the presence of type II like β-turns (Fig. 3)[26].

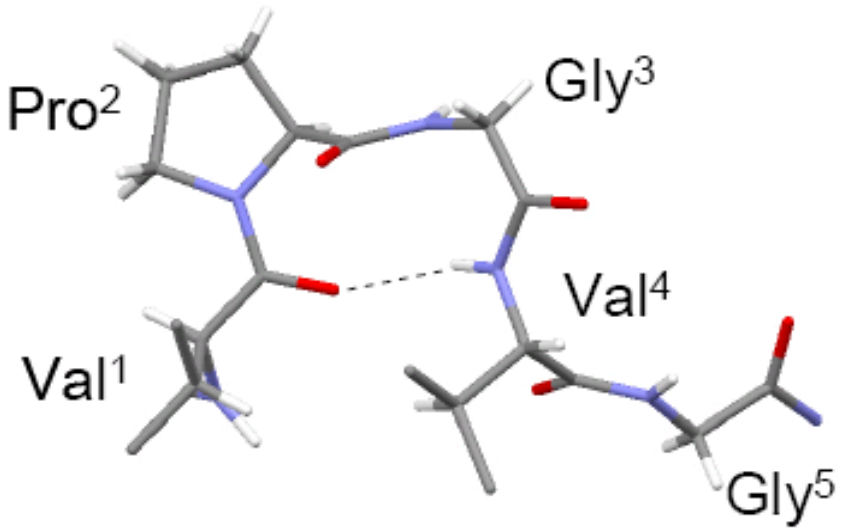


Figure 3. Type II β-turn structure of Pro^2-Gly^3 sequence motif showing a hydrogen bond between the Val^4 N-H and the Val^1 C=O obtained from the crystal structure of cyclo-$(VPGVG)_3$

In vivo, following secretion of tropoelastin into the extracellular matrix of target tissues, inter and intra-molecular cross-linking between lysine residues is carried out rapidly in the extracellular matrix by the copper-dependent enzyme, lysyl oxidase[32,33]. The enzyme targets most Lys residues in the tropoelastin sequence for cross-link formation. The enzyme-mediated cross-linking reaction involves oxidative deamination of a Lys residue to form allysine. Subsequent spontaneous condensation reactions follow

to link allysine and nearby lysine residues and form cross-links such as allysine aldol, lysinonorleucine, merodesmosine, and unique tetrafunctional cross-links such as desmosine and isodesmosine[3,34]. Cross-linking of tropoelastin results in the formation of mature, insoluble native elastin. Lysyl oxidase is responsible for forming Lys cross-links in native collagen as well[33,34]. Extraction and purification of tropoelastin from animal sources involves deprivation of copper from the diet of the animal leading to a loss in activity of lysyl oxidase and accumulation of tropoelastin in animal tissues[35,36]. Tropoelastin is one of many natural materials of interest in biomaterials engineering applications.

Collagen, for example, is a highly evolved macromolecule that provides tensile strength for the tissues of higher organisms as a major component of the extracellular matrix. Collagen is also the most abundant protein in animals and constitutes approximately one-third of the human proteome[37]. Collagen is an interesting biomaterial as the strength of the triple-helical, structural protein is controlled by the unique amino acid sequences of the polypeptides found in mature collagen. The structures of both native collagen and synthetic collagen analogues have been characterized by X-ray diffraction and other methods since the first structural model was presented in the 1950's[38]. Collagen consists of three polypeptide molecules that self-assemble into a triple helix *in vivo* after secretion into the extracellular matrix of target mammalian tissues. The resultant tropocollagen is further modified by a variety of factors that increase the stability of the mature protein[38].

The amino acid sequence of animal collagen is characterized by a repeating tripeptide unit, Gly-X-Y, where X and Y are often occupied by the imino acids proline or

hydroxyproline. The stability of the triple helical, coiled-coil conformation significantly depends upon the residue type at specific positions within the primary sequence. For example, hydroxyproline increases the stability of native collagen and of collagen-mimetic polypeptides[39]. The relatively rare amino acid, hydroxyproline, is formed from post-translational modification of proline by the enzyme prolyl-4-hydroxylase[40]. Raines and coworkers have demonstrated the self-assembly of collagen-mimetic proteins[37]. The collagen mimetics were synthesized by solid-phase peptide synthesis and self-assembly was carried out *in vitro*. The group has also provided valuable insight into the structure of collagen by conducting a series of studies on the effects that hydroxyproline and proline analogue residues have on the structure of synthetic collagen triple helices[37,38,41-47]. Large-scale methods for production of synthetic collagen for biomaterials applications are still under investigation as the insolubility and requirement for hydroxyproline incorporation present a significant challenge for synthetic methods. Biosynthesis of collagen is an attractive approach as the co-translational incorporation of hydroxyproline has been demonstrated in recombinant collagen and elastin-mimetic polypeptides[48-51].

Genetic Engineering of Biomaterials

Native elastin and collagen are both insoluble structural proteins that have been widely investigated for the purposes of biomaterials design and engineering[1]. Mature collagen is purified from animal sources for medical applications, but the insolubility of native elastin has restricted its capacity to be purified and processed into forms suitable for biomedical or industrial applications. Collagen is mainly isolated from bovine sources, but extraction of the native, insoluble biomaterial can present problems in bioengineering and utility of the material in medical applications. Recombinant protein

engineering can significantly increase protein yield over that which can be achieved by extraction of a native protein from animal tissues and offers the ability to use human amino acid sequences, so as to avoid adverse immunological responses and batch-to-batch variations.

Genetic engineering techniques have been very useful in the biosynthesis and high-level expression of a variety of elastin-mimetic polypeptides[52,53]. Generating polypeptide analogues, which retain the useful properties of native collagen and elastin through a biosynthetic approach such as protein engineering provides sequence control, minimizes batch-to-batch variation, and avoids contamination from pathogens. Longer protein polymers can be produced on a large scale as compared to synthesis of materials by solid-phase peptide synthesis. Biosynthetic methods potentially provide freedom from the necessity for post-translational modification of the synthetic protein *in vivo*[54]. For example, the co-translational, global incorporation of *trans*-4-hydroxyproline into elastin-mimetic proteins has been demonstrated in *E. coli* with protein yields on the order of 15 to 50 mg protein/L of culture[49]. In summary, recombinant protein engineering at the molecular (DNA) level is a useful technique for generating large amounts of protein polymers with versatile control over sequence specificity.

Hypothesis

If a gene is designed and synthesized to encode a novel elastin-mimetic protein, then expression of the protein (**elastin-CCC**) will be used as a test substrate for multiple, site-specific proline analogue incorporation in a bacterial host.

Directed Evolution

Design and genetic engineering of novel, orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pairs for multi-site incorporation of proline analogues in *E. coli*

Hydroxyproline residues present in native collagen are formed through post-translational hydroxylation of proline residues by the enzyme, prolyl hydroxylase[38]. Therefore, recent efforts to express recombinant collagen-based proteins have involved the coordinated co-expression of both a collagen target and prolyl-4-hydroxylase[40,55]. While some success has been achieved, low protein yields, and inhomogeneity of molecular structure due to inconsistent post-translational modification remain significant limitations of this approach. Collagen is characterized by the repeating tripeptide unit, Gly-X-Y, where X and Y are often occupied by the imino acids proline and hydroxyproline, respectively. Natural systems are able to hydroxylate proline residues at the third position, while adjacent prolines remain unmodified and serve to stabilize the triple helix. Co-expression of prolyl-4-hydroxylase alone does not allow for the sequence specificity required for biosynthesis of collagen or elastin-mimetics for biomaterial applications. In addition, the capacity to enhance collagen stability and/or functionality through the incorporation of alternate proline or hydroxyproline analogues is not feasible using this strategy.

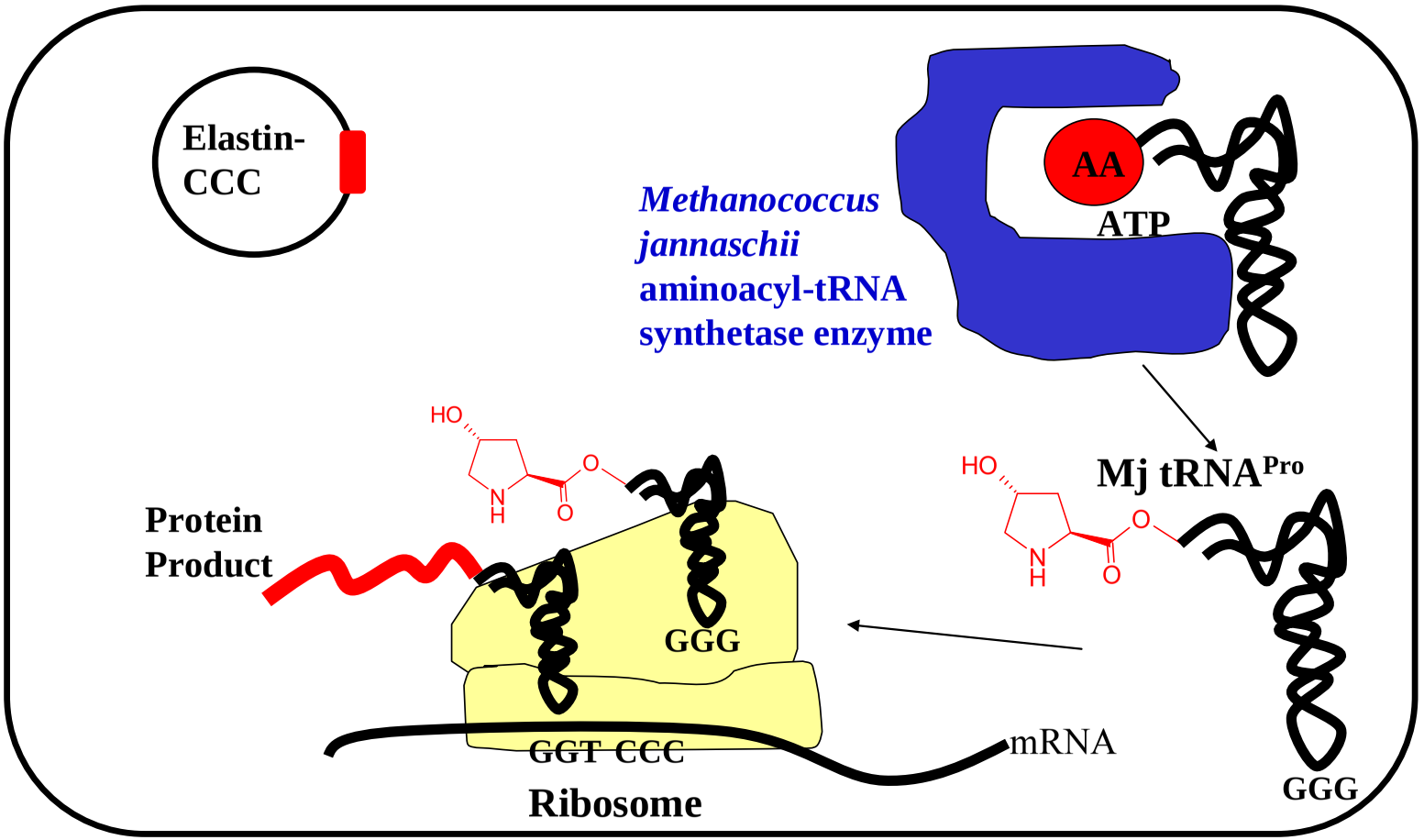
A potential solution to these problems may be found in the recent development of new experimental techniques that facilitate the direct incorporation of hydroxyproline or other proline analogues through an expansion of the genetic code[56]. In general terms, methods to expand the genetic code have included the selection and directed evolution of *orthogonal* tRNA/aaRS pairs that function in a heterologous host, such that a non-

canonical amino acid[57] is inserted in response to a termination codon using non-sense suppression[58-60]. An orthogonal tRNA/synthetase pair is introduced into the host using genetic engineering techniques, and the pair functions independently in the host organism as long as cross-species aminoacylation is inefficient and the anticodon loop is not a key determinant of synthetase/tRNA recognition. The synthetase enzymes catalyze aminoacylation of a cognate tRNA (with the appropriate anticodon) with the natural substrate amino acid or an amino acid analogue for incorporation into proteins during translation at the site of the ribosome (Scheme 1). While successful incorporation of non-canonical amino acids has been achieved, non-sense suppression has limited efficiency, particularly for multi-site incorporation of amino acid analogues. Therefore, a different approach was chosen that involved capture of codons that are rarely utilized in the native host organism. For example, codon usage statistics in *E. coli* suggest that a subset of the canonical codons occur very rarely in structural genes within the genome[61]. These codons are good candidates for capture using an orthogonal tRNA/aaRS system, as their use should not induce termination or significantly disrupt the normal function of the host organism.

Given this framework, the CCC codon of the proline family box was chosen as a target for incorporation of non-canonical proline analogues in an *E. coli* expression system. This approach was based upon the observations that mutant strains of the related organism *S. typhimurium* containing a dysfunctional version[62] or deletion[63] of the gene encoding the native tRNAPro(GGG) are viable. Furthermore, the CCC codon occurs at a very low frequency within structural genes in the *E. coli* genome, selective methods have been developed for knocking out specific tRNA genes in the *E. coli* genome[64], and

orthogonal tRNAPro/ProRS pairs exist in *H. sapiens*[65] and *M. jannaschii*[66]. An archaeal, *M. jannaschii* tRNA/synthetase pair, compared to the *E. coli* host, exhibits decreased sequence similarity to the host machinery, and one would expect to see different synthetase recognition elements between the two species[66]. An orthogonal pair would enable co-translational incorporation of proline analogues, such as (2*S*,4*R*)-4-hydroxyproline or (2*S*,4*R*)-4-fluoroproline, specifically in response to the CCC codon while canonical proline incorporation would be achieved in response to the other codons of the proline family box (Scheme 1)[67]. The experimental target systems involve the development of mutant versions of the *M. jannaschii* prolyl-tRNA synthetase enzyme that can utilize selected proline analogues, such as (2*S*,4*R*)-4-hydroxyproline, in preference to the canonical substrate imino acid. An elastin-mimetic model system where proline is encoded by the CCC codon for multiple, site-specific incorporation of proline analogues using the orthogonal system is ideal as a test system for a variety of reasons.

Elastin-mimetic polypeptides have been designed and produced in high yield in bacterial host systems[7], incorporation of a variety of proline analogues has been achieved in elastin-mimetics[49-51,68], and genetic engineering techniques have been established for the efficient and rapid assembly of synthetic genes encoding elastin-mimetic, repetitive polypeptides for direct cloning into bacterial expression vectors by an extension of the seamless cloning technique[27]. Proline is a major constituent of the elastin-mimetics, as it is in collagen, occupying ~20 % of the overall amino acid composition of elastin-mimetic polypeptides[49].



Scheme 1. Outline for using an orthogonal system for incorporation of (*2S,4R*)-4-hydroxyproline in an *E. coli* bacterial host. The **elastin-CCC** gene is encoded by an expression vector present in the host cell. An aminoacyl-tRNA synthetase/tRNA pair encoded by a separate expression plasmid to enable incorporation of hydroxyproline in **elastin-CCC** in response to the CCC codon upon induction. Other native amino acids are incorporated using the cellular machinery in the host.

Previous studies have also demonstrated the significant effects incorporation of proline analogues has on the stability and structure of both collagen-mimetics[45,46] and elastin-mimetic[50,51] proteins. Collagen alone as a test substrate for evolution of the orthogonal pair poses limitations such as the insolubility of collagen without modification of the molecular structure and lack of previous utility in experimental *in vivo* systems. Site-specific incorporation of proline analogues in elastin-mimetic protein polymers in response to specific codons will serve as a model for and enable the bacterial biosynthesis of synthetic collagen analogues and collagen-mimetic oligopeptides in high yield and at a significantly reduced cost.

Hypothesis

If a MjProRS/tRNA^{Pro} pair is constructed, using genetic engineering techniques, then the pair will function independently (in an orthogonal manner) in *E. coli* to enable incorporation of proline analogues in elastin-mimetic proteins.

Screening methodologies for engineering site-specificity in orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pairs for multi-site incorporation of proline analogues in *E. coli*

Several approaches are available for engineering altered substrate specificity in enzymes[69-76]. Directed evolution techniques have shown recent promise in engineering modified enzymes for various applications[70,73,74,77-79]. Aminoacyl-tRNA synthetase enzymes have been targeted in directed evolution studies and they have been successfully altered for use in orthogonal pairs for incorporation of amino acid analogues in bacterial systems[56,59,67,80-84]. For example, Peter Schultz and coworkers optimized a useful method for the evolution of aminoacyl-tRNA synthetase pairs[85]. The group was able to

selectively incorporate a series of tyrosine analogues into proteins in *E. coli* after generation of *M. jannaschii* tyrosyl-tRNA synthetase (MjYRS) mutant enzymes specific for the amino acid analogues. The mutant enzyme variants were used with a cognate tRNA as an orthogonal pair in the host for site-specific introduction of *p*-isopropyl-phenylalanine, *p*-amino-phenylalanine, *p*-carboxyl-phenylalanine, and *O*-allyl-phenylalanine. The side-chain binding pocket of the MjYRS enzyme exhibited flexibility in the accommodation of the different functional groups introduced and had been shown previously to selectively accept substrates other than tyrosine[81]. The strategy for generation of synthetase mutants that showed a preference for tyrosine analogues was based on high-throughput screening of large libraries of synthetase variants using reporter detection with fluorescence-activated cell sorting (FACS)[85].

FACS has been used widely in the fields of medicine and immunology for a wide variety of applications. FACS is often used to monitor the progression of diseases such as HIV/AIDS and cancer[86]. Using FACS, one is able to sort heterogeneous mixtures of cells into separate populations using a fluorescent reporter produced in or presented on the surface of the cells (Fig. 4)[86]. The majority of research involving FACS has been focused on optimization of the sorting technique with eukaryotic cells. The use of FACS presents a challenge with a goal for sorting bacterial cells as the cells are not only a different shape (rod-shaped in the case of *E. coli*) compared to roughly spherical eukaryotic cells, but also the diameter of a bacterial cell (1-2 μm) is much smaller that that of an average eukaryotic cell (10-20 μm)[87]. The Schultz group successfully used FACS in rapidly screening large libraries of synthetase variants using synthetase-dependent suppression of amber stop codons with the fluorescent reporter protein,

GFPuv[85]. Amber codons within the T7 RNA polymerase gene were recognized by the orthogonal synthetase/tRNA pairs to enable expression of GFPuv in cells where the synthetase variants were active with the analogue of interest, and bacterial cells expressing GFPuv were sorted using FACS in a positive/negative screening strategy alternating with antibiotic selection (chloramphenicol-resistance reporter). Cells harbored a plasmid with a chloramphenicol resistance gene with an amber stop codon position in the DNA sequence of the gene such that read-though of the gene using the orthogonal pair enabled chloramphenicol resistance. Cells expressing mutant synthetase enzymes that showed activity toward the tyrosine analogues could be selected directly using antibiotic resistance.

Directed evolution of the synthetase variants began with a mutagenesis strategy. The crystal structure of a homologous enzyme from *Bacillus stearothermophilus* was examined to determine the residues that made the closest contacts with the substrate in the synthetase enzyme. Five residues were chosen for site-saturation mutagenesis to afford a total library size of 3.2×10^6 possible permutations of the gene. The mutant libraries were subjected to several rounds of positive/negative screening and antibiotic selection to generate synthetase variants that no longer recognized the natural substrate, tyrosine, and were active toward the tyrosine analogues of interest *in vivo*. The group expressed proteins containing the non-canonical amino acids in *E. coli* on a mg protein/L culture scale[85].

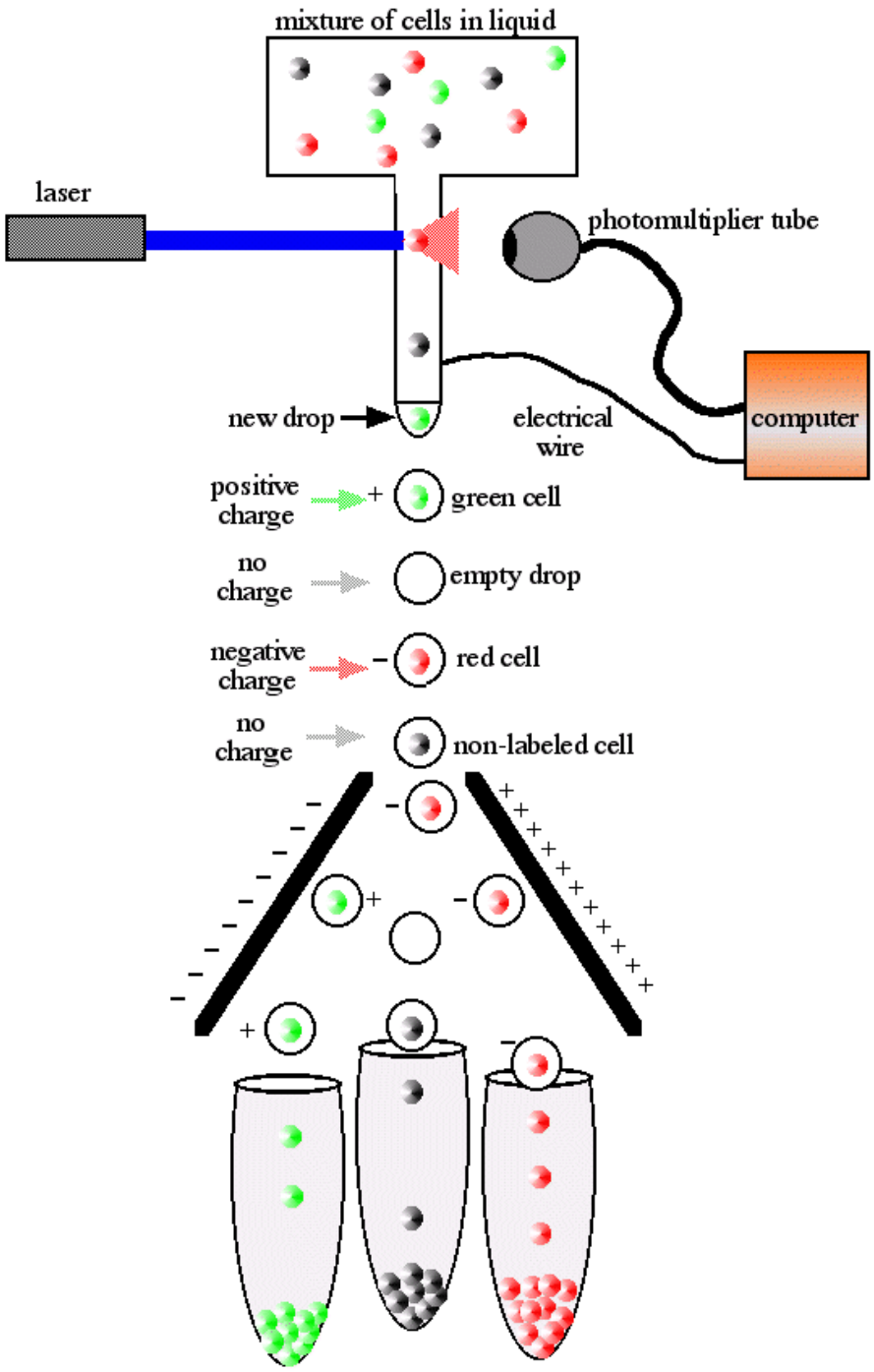


Figure 4. Fluorescence-activated cell sorting (FACS) enables isolation of separate cell populations expressing fluorescent reporter(s) from heterogeneous mixtures of cells. Separate droplets containing cells expressing a fluorescent reporter are detected and assigned a charge. As the charged droplets leave the fluid stream, they are deflected, based on their charge, by the electrical plates into separate collection chambers. The cells remain viable and are then used for further applications. Cells are often sorted in phosphate-buffered saline (PBS) with up to 98-99 % purity in each population sorted.

David Tirrell and coworkers also used FACS as a method for screening large libraries of mutant synthetase enzymes to generate novel methionyl-tRNA synthetases that enable residue-specific incorporation of non-canonical amino acids in proteins in a bacterial host[88]. A mutant GFP (GFPrm_AM) protein reporter was generated by site-saturation mutagenesis that exhibited fluorescence in the presence of the amino acid analogues of interest. As there were only two Met positions in the GFP variant, five additional positions in the primary sequence were changed to Met in order to generate the GFP used as a reporter for analogue incorporation. Four positions within the EcMetRS were chosen for site-saturation mutagenesis as they make close contacts with the natural substrate based on crystal structure data. Several rounds of positive/negative screening (3 positive, 1 negative) were conducted using FACS to enrich for synthetase variants with a preference for 6,6,6-trifluoronorleucine (Tfn) over the natural methionine substrate. Although the *in vitro* kinetic analysis of a mutant synthetase isolated from the screens indicated a sevenfold lower activation of Tfn compared to Met, cells grown in media depleted of Met were able to utilize and incorporate Tfn in GFP using the orthogonal pair. For strains with the mutant MetRS, the yields of proteins synthesized in Tfn-supplemented media were 20-30 mg protein/L culture[88].

Screening methodologies similar to those developed by the Tirrell and Schultz groups were investigated for engineering site-specificity in orthogonal tRNA/aminoacyl-tRNA synthetase ($tRNA^{Pro}$/ProRS) pairs for multi-site incorporation of proline analogues in *E. coli.* Both of the previously described studies enabled site-specific incorporation of amino acid analogues using an orthogonal pair, but neither study addressed high-level multi-site incorporation of proline analogues in protein-based materials. A few sites (1-

15) at most were replaced by the analogues of interest. Rational, site-saturation mutagenesis of the MjProRS enzyme was a promising first step toward isolation of a mutant MjProRS with specificity for (2*S*,4*R*)-4-hydroxyproline over the natural proline substrate. FACS-based screening approaches were shown to be useful for high-throughput screening of large libraries of synthetase mutants, so the technique was chosen for screening the MjProRS mutant library. Chapters 2-4 provide detail concerning the construction and evaluation of the orthogonal system, the mutant libraries, and of the reporter systems used in FACS-based screening of MjProRS variants for altered substrate specificity toward (2*S*,4*R*)-4-hydroxyproline. Generation of a pair with specificity for hydroxyproline over the natural substrate, proline, would be a significant step forward in the development of a route for synthetic collagen biosynthesis in bacteria. Site-specific incorporation of amino acid analogues in elastin-mimetic polypeptide materials would be an advance in the ability to generate protein polymers with ~20 % non-canonical amino acid content overall compared to incorporation into a few selected sites in catalytic proteins. The development of a screening system for generating a novel, orthogonal tRNA/aaRS pair alone is a significant advance in the fields of directed evolution and biomaterials engineering. The screening system and pair would be highly valuable tools in the molecular toolbox available for chemists and biomedical engineers in the development and production of novel protein-based biomaterials.

Hypothesis

If a directed evolution and FACS-based screening strategy is developed, then the screen will be used for the isolation of MjProRS variants that demonstrate a preference for (2*S*,4*R*)-4-hydroxyproline over the native substrate, proline.

References

(1) van Hest, J. C. M.; Tirrell, D. A. *Chemical Communications* **2001**, 1897-1904.

(2) Mithieux, S. M.; Weiss, A. S. In *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers* 2005; Vol. 70, p 437-+.

(3) Vrhovski, B.; Weiss, A. S. *European Journal of Biochemistry* **1998**, *258*, 1-18.

(4) Lillie, M. A.; Gosline, J. M. *International Journal of Biological Macromolecules* **2002**, *30*, 119-127.

(5) Gosline, J.; Lillie, M.; Carrington, E.; Guerette, P.; Ortlepp, C.; Savage, K. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **2002**, *357*, 121-132.

(6) Powell, J. T.; Vine, N.; Crossman, M. *Atherosclerosis* **1992**, *97*, 201-208.

(7) Daamen, W. F.; Veerkamp, J. H.; van Hest, J. C. M.; van Kuppevelt, T. H. *Biomaterials* **2007**, *28*, 4378-4398.

(8) Olliver, L.; Luvalle, P. A.; Davidson, J. M.; Rosenbloom, J.; Mathew, C. G.; Bester, A. J.; Boyd, C. D. *Collagen and Related Research* **1987**, *7*, 77-89.

(9) Indik, Z.; Yeh, H.; Ornsteingoldstein, N.; Sheppard, P.; Anderson, N.; Rosenbloom, J. C.; Peltonen, L.; Rosenbloom, J. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 5680-5684.

(10) Bashir, M. M.; Indik, Z.; Yeh, H.; Ornsteingoldstein, N.; Rosenbloom, J. C.; Abrams, W.; Fazio, M.; Uitto, J.; Rosenbloom, J. *Journal of Biological Chemistry* **1989**, *264*, 8887-8891.

(11) Boyd, C. D.; Christiano, A. M.; Pierce, R. A.; Stolle, C. A.; Deak, S. B. *Matrix* **1991**, *11*, 235-241.

(12) Fazio, M. J.; Mattei, M. G.; Passage, E.; Chu, M. L.; Black, D.; Solomon, E.; Davidson, J. M.; Uitto, J. *American Journal of Human Genetics* **1991**, *48*, 696-703.

(13) Indik, Z.; Yoon, K.; Morrow, S. D.; Cicila, G.; Rosenbloom, J.; Ornsteingoldstein, N. *Connective Tissue Research* **1987**, *16*, 197-211.

(14) Urry, D. W. *Journal of Physical Chemistry B* **1997**, *101*, 11007-11028.

(15) Rucker, R. B.; Dubick, M. A. *Environmental Health Perspectives* **1984**, *55*, 179-191.

(16) Urry, D. W. *Journal of Protein Chemistry* **1988**, *7*, 1-34.

(17) Urry, D. W. *Perspectives in Biology and Medicine* **1978**, *21*, 265-295.

(18) Urry, D. W. *Scientific American* **1995**, *272*, 64-69.

(19) Wise, S. G.; Mithieux, S. M.; Raftery, M. J.; Weiss, A. S. *Journal of Structural Biology* **2005**, *149*, 273-281.

(20) Rapaka, R. S.; Okamoto, K.; Urry, D. W. *International Journal of Peptide and Protein Research* **1978**, *12*, 81-92.

(21) Urry, D. W.; Mitchell, L. W.; Ohnishi, T. *Biochimica Et Biophysica Acta* **1975**, *393*, 296-306.

(22) Urry, D. W.; Harris, R. D.; Long, M. M.; Prasad, K. U. *International Journal of Peptide and Protein Research* **1986**, *28*, 649-660.

(23) Gray, W. R.; Sandberg, L. B.; Foster, J. A. *Nature* **1973**, *246*, 461-466.

(24) Urry, D. W.; Long, M. M.; Cox, B. A.; Ohnishi, T.; Mitchell, L. W.; Jacobs, M. *Biochimica Et Biophysica Acta* **1974**, *371*, 597-602.

(25) Cook, W. J.; Einspahr, H.; Trapane, T. L.; Urry, D. W.; Bugg, C. E. *Journal of the American Chemical Society* **1980**, *102*, 5502-5505.

(26) Urry, D. W.; Luan, C. H.; Peng, S. Q. In *Molecular Biology and Pathology of Elastic Tissues* 1995; Vol. 192, p 4-30.

(27) McMillan, R. A.; Lee, T. A. T.; Conticello, V. P. *Macromolecules* **1999**, *32*, 3643-3648.

(28) McMillan, R. A.; Conticello, V. P. *Macromolecules* **2000**, *33*, 4809-4821.

(29) McPherson, D. T.; Morrow, C.; Minehan, D. S.; Wu, J. G.; Hunter, E.; Urry, D. W. *Biotechnology Progress* **1992**, *8*, 347-352.

(30) Panitch, A.; Yamaoka, T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1999**, *32*, 1701-1703.

(31) Trabbic-Carlson, K.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2003**, *4*, 572-580.

(32) Bressan, G. M.; Prockop, D. J. *Biochemistry* **1977**, *16*, 1406-1412.

(33) Kagan, H. M.; Sullivan, K. A. *Methods in Enzymology* **1982**, *82*, 637-650.

(34) Reiser, K.; McCormick, R. J.; Rucker, R. B. *Faseb Journal* **1992**, *6*, 2439-2449.

(35) Rucker, R. B. *Methods in Enzymology* **1982**, *82*, 650-657.

(36) Sandberg, L. B.; Wolt, T. B. *Methods in Enzymology* **1982**, *82*, 657-665.

(37) Kotch, F. W.; Raines, R. T. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 3028-3033.

(38) Jenkins, C. L.; Raines, R. T. *Natural Product Reports* **2002**, *19*, 49-59.

(39) Raines, R. T. *Protein Science* **2006**, *15*, 1219-1225.

(40) Kersteen, E. A.; Higgin, J. J.; Raines, R. T. *Protein Expression and Purification* **2004**, *38*, 279-291.

(41) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666-667.

(42) Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. *Chemistry & Biology* **1999**, *6*, 63-70.

(43) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. *Journal of the American Chemical Society* **2001**, *123*, 777-778.

(44) Hodges, J. A.; Raines, R. T. *Biochemistry* **2001**, *40*, 8657-8657.

(45) Hodges, J. A.; Raines, R. T. *Journal of the American Chemical Society* **2003**, *125*, 9262-9263.

(46) Jenkins, C. L.; Bretscher, L. E.; Guzei, I. A.; Raines, R. T. *Journal of the American Chemical Society* **2003**, *125*, 6422-6427.

(47) Kotch, F. W.; Raines, R. T. *Biopolymers* **2005**, *80*, 594-595.

(48) Buechter, D. D.; Paolella, D. N.; Leslie, B. S.; Brown, M. S.; Mehos, K. A.; Gruskin, E. A. *Journal of Biological Chemistry* **2003**, *278*, 645-650.

(49) Kim, W. Y.; George, A.; Evans, M.; Conticello, V. P. *Chembiochem* **2004**, *5*, 928-936.

(50) Kim, W.; McMillan, R. A.; Snyder, J. P.; Conticello, V. P. *Journal of the American Chemical Society* **2005**, *127*, 18121-18132.

(51) Kim, W.; Hardcastle, K. I.; Conticello, V. P. *Angewandte Chemie-International Edition* **2006**, *45*, 8141-8145.

(52) Rodriguez-Cabello, J. C.; Reguera, J.; Girotti, A.; Alonso, M.; Testera, A. M. *Progress in Polymer Science* **2005**, *30*, 1119-1145.

(53) Tirrell, D. A.; Fournier, M. J.; Mason, T. L. *Mrs Bulletin* **1991**, *16*, 23-28.

(54) Connor, R. E.; Tirrell, D. A. *Polymer Reviews* **2007**, *47*, 9-28.

(55) Olsen, D.; Yang, C. L.; Bodo, M.; Chang, R.; Leigh, S.; Baez, J.; Carmichael, D.; Perala, M.; Hamalainen, E. R.; Jarvinen, M.; Polarek, J. *Advanced Drug Delivery Reviews* **2003**, *55*, 1547-1567.

(56) Wang, L.; Schultz, P. G. *Angewandte Chemie-International Edition* **2005**, *44*, 34-66.

(57) Link, A. J.; Mock, M. L.; Tirrell, D. A. *Current Opinion in Biotechnology* **2003**, *14*, 603-609.

(58) Santoro, S. W.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 4185-4190.

(59) Santoro, S. W.; Anderson, J. C.; Lakshman, V.; Schultz, P. G. *Nucleic Acids Research* **2003**, *31*, 6700-6709.

(60) Santoro, S.; Schultz, P. *Directed evolution of the substrate specificities of a site-specific recombinase and an aminoacyl-tRNA synthetase using fluorescence-activated cell sorting (FACS). Methods in Molecular Biology.* 2003; Vol. 230.

(61) Kane, J. F. *Current Opinion in Biotechnology* **1995**, *6*, 494-500.

(62) Qian, Q.; Bjork, G. R. *Journal of Molecular Biology* **1997**, *273*, 978-992.

(63) Chen, P.; Qian, Q.; Zhang, S. P.; Isaksson, L. A.; Bjork, G. R. *Journal of Molecular Biology* **2002**, *317*, 481-492.

(64) McClain, W. H.; Gabriel, K. *Journal of Molecular Biology* **2001**, *310*, 537-542.

(65) Stehlin, C.; Burke, B.; Yang, F.; Liu, H. J.; Shiba, K.; Musier-Forsyth, K. *Biochemistry* **1998**, *37*, 8605-8613.

(66) Beuning, P. J.; Musier-Forsyth, K. *Journal of Biological Chemistry* **2001**, *276*, 30779-30785.

(67) Kwon, I.; Kirshenbaum, K.; Tirrell, D. A. *Journal of the American Chemical Society* **2003**, *125*, 7512-7513.

(68) Kim, W.; Conticello, V. P. *Polymer Reviews* **2007**, *47*, 93-119.

(69) Farinas, E. T.; Bulter, T.; Arnold, F. H. *Current Opinion in Biotechnology* **2001**, *12*, 545-551.

(70) Arnold, F. H.; Wintrode, P. L.; Miyazaki, K.; Gershenson, A. *Trends in Biochemical Sciences* **2001**, *26*, 100-106.

(71) Lin, H. N.; Cornish, V. W. *Angewandte Chemie-International Edition* **2002**, *41*, 4403-4425.

(72) Lutz, S.; Patrick, W. M. *Current Opinion in Biotechnology* **2004**, *15*, 291-297.

(73) Bloom, J. D.; Meyer, M. M.; Meinhold, P.; Otey, C. R.; MacMillan, D.; Arnold, F. H. *Current Opinion in Structural Biology* **2005**, *15*, 447-452.

(74) Montclare, J. K.; Tirrell, D. A. *Angewandte Chemie-International Edition* **2006**, *45*, 4518-4521.

(75) Haseltine, E. L.; Arnold, F. H. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36*, 1-19.

(76) Toscano, M. D.; Woycechowsky, K. J.; Hilvert, D. *Angewandte Chemie-International Edition* **2007**, *46*, 3212-3236.

(77) Wang, L.; Xie, J.; Schultz, P. G. *Annual Review of Biophysics and Biomolecular Structure* **2006**, *35*, 225-249.

(78) Ryu, Y. H.; Schultz, P. G. *Nature Methods* **2006**, *3*, 263-265.

(79) Chen, S.; Schultz, P. G.; Brock, A. *Journal of Molecular Biology* **2007**, *371*, 112-122.

(80) Wang, L.; Schultz, P. G. *Chemical Communications* **2002**, 1-11.

(81) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

(82) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Journal of the American Chemical Society* **2002**, *124*, 9026-9027.

(83) Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G. *Journal of the American Chemical Society* **2003**, *125*, 935-939.

(84) Xie, J. M.; Schultz, P. G. *Methods* **2005**, *36*, 227-238.

(85) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nature Biotechnology* **2002**, *20*, 1044-1048.

(86) Shapiro, H. M. *Practical Flow Cytometry*; 4th ed.; John Wiley and Sons, Inc.: Hoboken, NJ, 2003.

(87) Garrett, R. H., Grisham, C.M. *Biochemistry*; 2nd ed.; Thomson, 1999.

(88) Yoo, T. H.; Tirrell, D. A. *Angewandte Chemie-International Edition* **2007**, *46*, 5340-5343.

Chapter 2

Generation of a Synthetic Gene Encoding an Elastin-Mimetic Polypeptide, Elastin-CCC

Introduction

The production of protein-based polymers is a useful approach for biomaterials design[1-5]. Our lab has achieved success in the design and engineering of a variety of protein polymer materials[6-17]. Genetic engineering techniques are particularly useful for generating protein block copolymers as controlling the synthesis of a protein at the DNA level allows for high-level protein expression in a bacterial host[18,19]. One can generate target proteins with an *in vivo* approach that allows for biosynthesis of a protein with predetermined DNA and subsequent amino acid sequences. Controlling the sequence of the polypeptide enables production of target biomaterials with defined polymer size, sequence specificity, and a low error rate. Our lab has produced several protein block copolymer materials with desirable properties and structure-function relationships using genetic engineering techniques[6,8,9,11,12,20,21].

In order to synthesize large genes encoding tandem repeats, such as the repeat sequences of collagen and elastin, one must design an effective strategy for production of protein materials in the bacterial host. The strategy outlined in Figure 1 is a general scheme for production of protein polymers. An effective strategy for the biological synthesis of repetitive polypeptides involves chemical synthesis of a corresponding DNA coding sequence, ligation into a plasmid vector, propagation in a bacterial host, and inducible expression of the polypeptide. In order to synthesize large genes encoding tandem repeats of the target oligopeptide sequence, it is necessary to begin with assembly of DNA monomers containing non-palindromic, cohesive ends. The DNA monomers are

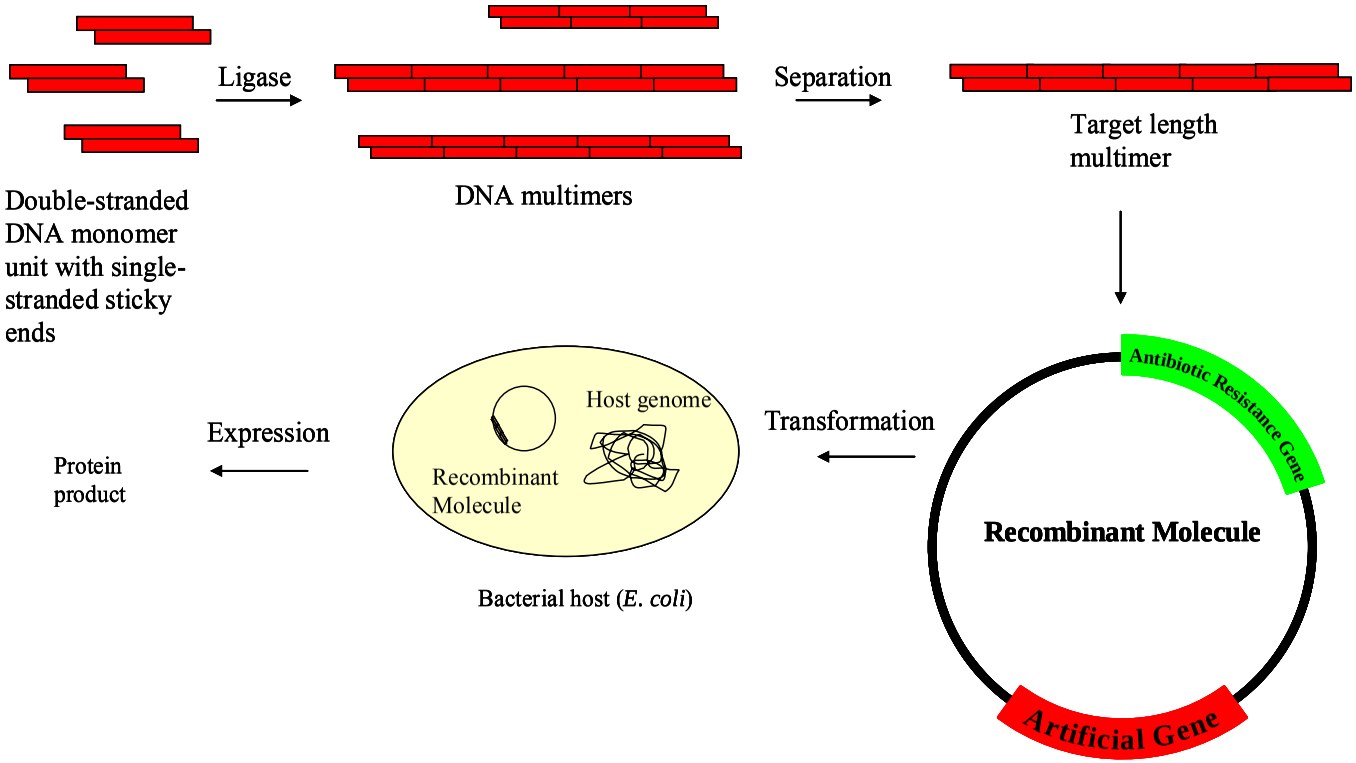


Figure 1. Synthesis of repetitive polypeptides via concatemerization of DNA cassettes

produced via annealing the single stranded DNA sequences together to produce the double-stranded monomers with cohesive ends. The DNA monomers are oligomerized together in a head-to-tail fashion such that the sense strand is translated into the polypeptide. The DNA concatemers are separated by size using agarose gel electrophoresis and enzymatically inserted into an expression vector. The vector is then transformed into an *E. coli* expression host. Production of the gene is controlled by a promoter contained in the vector. The promoter induces the expression of the polypeptide in the bacterial host. This method allows for the production of up to gram quantities of the desired protein[6]. The concatemerization technique was utilized for the assembly of a synthetic gene encoding an elastin-mimetic polypeptide based on the pentapeptide repeat sequence, Val-Pro-Gly-Val-Gly. The phase behavior and mechanical properties of elastin-mimetic biomaterials depend critically on the identity of the residues within the repeat sequence[12,22].

The goal for production of elastin-mimetic proteins in this particular case involved constructing a DNA monomer sequence for the elastin-mimetic gene in which all of the positions occupied by proline would be encoded by the CCC codon. Because genetic engineering at the DNA level allows for sequence specific synthesis of a protein of interest, a genetic engineering strategy was used to generate an elastin-mimetic protein target that utilized the codon bias inherent in the *E. coli* host. A codon capture strategy was chosen in which the rarely used CCC codon in *E. coli*, normally encoding L-proline, encodes for proline analogues (*trans*-4-hydroxyproline) or proline itself. The goal for production of elastin-mimetic proteins in this case involved constructing a DNA monomer sequence for the elastin-mimetic gene in which all of the positions occupied by

Elastin-CCC Monomer Sequence

Bbs I Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Ile Gly Val Pro Gly Val
5′-AGCTTGAAGACGTACCCGGTGTTGGCGTTCCCGGTGTAGGTGTTCCCGGCATCGGTGTACCCGGTGTT
3′-ACTTCTGCATGGGCCACAACCGCAAGGGCCACATCCACAAGGGCCGTAGCCACATGGGCCACAA

Gly Val Pro Gly Val Gly Val Pro *Bsm*B I
GGCGTACCCGGCGTTGGCGTACCCAGAGACGG-3′
CCGCATGGGCCGCAACCGCATGGGTCTCTGCCGATC-5′

Primers used for construction of the **elastin-CCC** monomer

Ile25CCC-F
5’-AGCTTGAAGACGTACCCGGTGTTGGCGTTCCCGGTGTAGGTGTTCCCGGCATCGGTGTACCCGGTGTTGGCGTACCC
GGCGTTGGCGTACCCAGAGACGG-3’

Ile25CCC-R
5’-GATCCCGTCTCTGGGTACGCCAACGCCGGGTACGCCAACACCGGGTACACCGATGCCGGGAACACCTACACCGGGA
ACGCCAACACCGGGTACGTCTTCA-3’

Figure 2. Coding sequence for **elastin-CCC** monomer DNA.

proline or a proline analogue of interest would be encoded by the CCC codon. Figure 2 depicts the **elastin-CCC** monomer coding sequence. E**lastin-CCC** will serve as a model system for biosynthesis of collagen-mimetic proteins. High-level, accurate, co-translational incorporation of proline analogues, such as *trans*-4-hydroxy-L-proline, in addition to incorporation of L-proline in the same polypeptide is one of the major future goals for the project. The **elastin-CCC** coding sequence is of particular importance in construction of the orthogonal system that would enable site-specific incorporation of hydroxyproline in elastin-mimetic materials.

Materials

All chemical reagents were purchased from either Fisher Scientific, Inc. (Pittsburgh, PA) or Sigma Chemical Co. (St. Louis, MO) unless otherwise noted. Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases, T4 DNA ligase, and T4 kinase were purchased from New England Biolabs, Inc. (Beverly, MA), shrimp alkaline phosphatase was obtained from Roche Applied Science (Indianapolis, IN), and Platinum *Pfx* DNA polymerase was obtained from Invitrogen Corp. (Carlsbad, CA). Plasmids pQE-60 and pQE-80L were obtained from QIAGEN, Inc. (Valencia, CA). Synthetic oligonucleotides were purchased from either Sigma-Genosys, Inc (The Woodlands, TX) or Integrated DNA Technologies (Coralville, IA) and were used as received. The *E. coli* strain, Top10F´ was obtained from Invitrogen Corp. (Carlsbad, CA). Reagents for the manipulation of bacteria and DNA were sterilized by either autoclave or passage through a 0.2 μm filter. Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. Site-directed mutagenesis was performed using Stratagene's (La Jolla,

CA) Quick-Change mutagenesis technique from gene-specific oligonucleotide primers. Automated DNA sequencing was performed at the Emory University Core DNA Sequencing Facility on a Perkin-Elmer ABI Prism model 377 DNA sequencer.

Methods

Common lab procedures were obtained from a standard molecular cloning manual[23] unless otherwise described in detail.

Construction of the Elastin-CCC gene

DNA oligonucleotide primers (Ile25CCC-F and Ile25CCC-R) encoding the sense and anti-sense **elastin-CCC** monomer sequence (Fig. 2) were chemically synthesized, and the single-stranded primers were annealed to produce duplex monomer DNA. Annealing was carried out by dissolving the primers in distilled, deionized water (ddH_2O) to a final concentration of 0.5 μg/μL. A 10 μL aliquot of each of the two primers was mixed together with 4 μL 5M NaCl, 4 μL $MgCl_2$, and 172 μL ddH_2O. By gradually decreasing the temperature of the reactions from 99 °C to 30 ºC (decreased by 5 °C every 3 min), the DNA strands were annealed together. The monomer DNA was phosphorylated at the 5′ and 3′ positions using a T4 polynucleotide kinase enzyme. Ethanol precipitation was used to purify the duplex DNA fragment. The double-stranded DNA fragment was visualized by DNA agarose gel electrophoresis (4 % NuSieve agarose). A ligation between the monomer DNA and *Bam*H I/*Hind* III-digested pZErO®-1 (Invitrogen) plasmid was performed in the presence of T4 DNA ligase enzyme with incubation at 16 ºC for 12 h. An aliquot (2 μL) of the ligation mixture was used to transform *E. coli* strain, Top10F′. The cells were recovered in 1 mL SOC rich media for 1h. An aliquot (100 μL) of the transformant suspension was spread onto solid low salt

LB media supplemented with zeocin™ (50 μg/mL) for antibiotic selection. The plates were incubated at 37 °C for 12-14 h.

Eight single colonies were picked from the plate and used to inoculate seven culture tubes with 5 mL low salt LB media with zeocin that were grown overnight at 37 ºC with shaking at 225 rpm. The plasmid DNA was isolated using the QIAprep-spin miniprep kit (QIAGEN, Inc.) for each of the samples. Recombinant plasmids were screened by double digestion with *Bam*H I/*Hind* III to identify the **elastin-CCC** gene in the plasmid. Agarose gel electrophoresis was used to visualize the DNA. The sequence of the **elastin-CCC** monomer gene was confirmed by automated DNA sequencing analysis using oligonucleotide primers designed to bind upstream and downstream of the target sequence. All of the plasmid DNA samples contained the correct gene of interest. The plasmid, pHC3, is comprised of the pZErO®-1 plasmid with the **elastin-CCC** monomer insert flanked by *Bbs* I and *Bsm*B I restriction enzyme sites. DNA was stored in solution at -20 ºC when not in use.

A large amount of the monomer was needed for DNA cassette concatemerization. Approximately 1 mg of pHC3 plasmid DNA was isolated using the QIAfilter plasmid maxi kit (QIAGEN, Inc.). The plasmid DNA was digested with *Bbs* I for 16 h at 37 ºC. The DNA was purified by extraction with phenol/chloroform/isoamyl alcohol (25/24/1) to remove proteins and ethanol precipitation was used to purify the DNA and reduce the salt concentration. The entire isolated DNA sample (240 μg) was digested with *Bsm*B I at 55 °C for 16 h. The DNA was purified by extraction with phenol/chloroform/isoamyl alcohol (25/24/1) and ethanol precipitation. The DNA was loaded onto a preparative 2% NuSieve agarose gel in order to separate the digested pZErO®-1 plasmid from the

elastin-CCC monomer. **Elastin-CCC** monomer DNA was extracted from the gel, macerated, and purified using the Amicon Ultrafree MC Maximum Recovery Kit (Millipore, Burlington, MA). The DNA was further purified by ethanol precipitation (DNA yield: 4.3 μg determined by 4 % agarose gel electrophoresis). A ligation between monomers (multimerization) was prepared by incubating the isolated monomer DNA (3.75 μg) in the presence of T4 DNA ligase at 16 °C for 16 h. The multimerization reaction (50 μL total, ~20 μL in each of 3 wells) was loaded onto a 1% agarose gel and flanked by DNA markers (Fig. 3). The DNA was separated to reveal a ladder of DNA multimers that represented concatemers molecules of the **elastin-CCC** monomer. The desired length for the **elastin-CCC** gene was between 2000 and 1200 bp. The region of the gel corresponding to the desired multimer size range was excised from the gel and DNA was isolated using the Zymogen gel extraction kit manufactured by Zymo Research (Orange, CA).

Multimers in the range of 1000-100 bp and 3000-2000 bp were also isolated from the gel. The isolated collection of DNA multimers was ligated back into the *Bbs* I digested pHC3 parent plasmid containing one copy of the **elastin-CCC** monomer in pZErO®-1. An aliquot of the ligation was transformed into Top10F´. Single colonies were selected and DNA, representing the **elastin-CCC** gene in pZErO®-1, was isolated using the automated MacConnell miniprep-24 system (MacConnell Research, San Diego, CA). A total of 48 DNA samples were isolated, digested with *Bam*H I/*Hind* III, and then screened by 1 % agarose gel electrophoresis for the presence of multimer inserts.

Fourteen colonies representing the most promising inserts were cultured, and DNA was isolated from the bacteria using the QIAprep spin miniprep kit. The DNA was

transformed into the Stratagene (La Jolla, CA) SURE cell line, and DNA was isolated from the bacteria using the QIAprep spin miniprep kit. DNA samples were then separated using 1 % agarose gel electrophoresis, and six samples were identified for target multimer inserts (two samples in the 2000 bp range, two samples in the 1500-1200 bp range, one sample in the 500 bp range, and one sample in the 200 bp range). The sequence of the **elastin-CCC** gene multimers was confirmed by automated DNA sequencing analysis using oligonucleotide primers designed to bind upstream and downstream of the target sequence. For the two samples in the 1500-1200 bp range, both displayed the correct gene sequence for the complete **elastin-CCC** gene. Two samples in the 2000-1800 bp range were sequenced, and both displayed the correct gene sequence for the complete **elastin-CCC** gene. Next, an acceptor plasmid suitable for protein expression was required to conduct high-level expression of **elastin-CCC** in a bacterial host.

Expression Plasmid Construction

It was necessary to remove a *Bsa* I restriction endonuclease site from the plasmid, pQE-60 for cloning of the **elastin-CCC** gene cassette. The internal *Bsa* I restriction site within the plasmid vector was removed via site-directed mutagenesis, which introduced a silent mutation into the sequence of the plasmid, producing plasmid pHC4. An adaptor sequence was designed for cloning of the **elastin-CCC** gene cassette in the pQE-60 plasmid. The adaptor sequence is shown in Figure 3.

The forward and reverse primers corresponding to the adaptor sequence were annealed using a BioRad MJ Mini thermocycler (Hercules, CA). Annealing was carried out by dissolving the primers in distilled, deionized water (ddH_2O) to a final

concentration of 0.5 μg/μL. Ten microliters of each of the two primers were mixed together with 4 μL 5M NaCl, 4 μL $MgCl_2$, and 172 μL ddH_2O. By gradually decreasing the temperature of the reactions from 99 °C to 30 ºC (decreased by 5 °C every 3 min), the DNA strands were annealed together. The double-stranded DNA fragment was visualized by DNA agarose gel electrophoresis (4 % NuSieve agarose). The adaptor DNA was phosphorylated at the 5′ and 3′ positions using a T4 polynucleotide kinase enzyme. The pHC4 plasmid was double-digested with *Nco* I/*Hind* III restriction endonuclease enzymes and the adaptor DNA was ligated into the pQE-60-based expression plasmid yielding plasmid pHC5. The sequence of the adaptor sequence was confirmed by automated DNA sequencing analysis using primers designed to bind to the pQE-60 plasmid upstream and downstream of the multiple cloning sites.

Cloning Elastin-CCC into the Expression Plasmid

Four pZErO®-1 plasmids each containing 1500 bp and 1800 bp (two of each) were sequentially digested with *Bbs* I and *Bsm*B I to isolate the **elastin-CCC** gene. The pHC5 (pQE-60) plasmid, containing the adaptor sequence, was digested with *Bsa* I to generate the complimentary sticky ends necessary for seamless cloning of the **elastin-CCC** genes into the adaptor sequence of the expression plasmid (Fig. 3). The digested pHC5 DNA was also dephosphorylated with shrimp alkaline phosphatase to reduce the incidence of false positive ligations. Ligation reactions were incubated at 16 °C for 12 h to covalently link the digested pHC5 DNA, and two of the digested, **elastin-CCC** multimers (one at ~1500 bp and one at ~1800 bp). An aliquot of each of the two ligations was transformed in Top10F′ and eight single colonies were picked for screening. DNA was isolated from the cultured colonies under ampicillin (100 μg/mL) selection using the

Primers used for construction of the elastin-CCC adaptor sequence

CCCAdaptor-F
5´-CATGGTACCCAGAGACCGTCGACGGTCTCGACCCGGTGTTGGCGGATCCG
ACGACGACGACAAGGGCCATCATCATCATCATCATCATCATCACTAATA-3´

CCCAdaptor-R
5´-AGCTTATTAGTGATGATGATGATGATGATGATGATGATGGCCCTTGTCGTC
GTCGTCGGATCCGCCAACACCGGGTCGAGACCGTCGACGGTCTCTGGGTAC-3´

Adaptor sequence after annealing

Nco I *Bsa* I *Bsa* I *Bam*H I
5´-CATGGTACCCAGAGACCGTCGACGGTCTCGACCCGGTGTTGGCGGATCCGACGACGACGACAAGGGCCAT
3´-ATGGGTCTCTGGCAGCTGCCAGAGCTGGGCCACAACCGCCTAGGCTGCTGCTGCTGTTCCCGGTA

CATCATCATCATCATCATCATCATCACTAATA-3´
GTAGTAGTA GTAGTAGTAGTAGTAGTGATTATTCGA-5´
Decahistidine Tag *Hin*d III

Figure 3. Elastin-CCC adaptor sequence.

QIAprep spin miniprep kit. The sequences of the **elastin-CCC** gene multimers were confirmed by automated DNA sequencing analysis using primers designed to bind to the pQE-60 plasmid upstream and downstream of the adaptor sequence. pHC6 includes the pQE-60 plasmid with the 1500 bp size **elastin-CCC** gene insert while the pHC7 plasmid includes the pQE-60 plasmid with the 1800 bp insert.

The pQE-80L plasmid provides an advantage over the use of the pQE-60 plasmid as the pQE-80L plasmid contains a copy of the overproducing repressor allele, $lacI_q$, to ensure tight control of the basal level of transcription prior to induction of **elastin-CCC** protein expression with IPTG. The pHC6, pHC7, and the pQE-80L plasmid DNA samples were then double digested with *Eco*R I/*Hind* III. The pQE-80L plasmid was dephosphorylated by incubation with shrimp alkaline phosphatase at 37 °C for 1 h. The DNA was gel purified by electrophoretic separation on a 1 % agarose gel and excision of DNA bands of correct size. Gel purification was carried out using the Zymo gel extraction kit (Zymo Research). The digested, multimer DNA fragments were cloned into the digested pQE-80L plasmid via ligation and aliquots of the ligations were transformed into Top10F´ and plated on solid LB media supplemented with ampicillin (100 µg/mL) for selection of recombinant multimers in pQE-80L. Single colonies were picked and cultured overnight in 5 mL LB/ampicillin (100 µg/mL). DNA was isolated with the QIAprep spin miniprep kit, and samples of DNA were sent for sequencing. The sequences of the **elastin-CCC** gene multimers were confirmed by automated DNA sequencing analysis using primers designed to bind to the pQE-60 plasmid upstream and downstream of the gene sequence. Both plasmid pHC8, containing the ~1500 bp insert, and pHC9, containing the ~1800 bp insert, had the correct sequence for the **elastin-CCC**

gene present in pQE-80L. Frozen stocks were prepared for each of the plasmids constructed by combining 800 μL of a culture of bacteria, containing the plasmid to be stored, grown to an OD_{600} of ~0.6 and 200 μL of 80 % glycerol. Frozen stocks were stored at -80ºC.

Results and Discussion

The repetitive domain of **elastin-CCC** comprises a concatenated series of pentapeptide repeats based on the consensus sequence (Val-Pro-Gly-Val/Ile-Gly) in which proline residues constitute 20 % of the amino acids within the repetitive domain (80 prolines/polypeptide chain). The high density of proline residues and the uniform structural environment of these residues within the polypeptide sequence provides for a well-defined context for evaluating the efficacy of analogue incorporation in terms of its effect on protein yield and analogue substitution level. Biophysical studies of elastin and elastin-mimetic polypeptides have suggested a crucial structural role for proline residues within the pentapeptide repeats that may be essential for the normal physiological function of the native polypeptide material. The concatemerization technique was successfully used to generate a large, full-length gene encoding the **elastin-CCC** protein. The results of multimerization are shown in Figure 4. In addition, several **elastin-CCC** genes of variable lengths were generated that would be useful for further studies (Fig. 10).

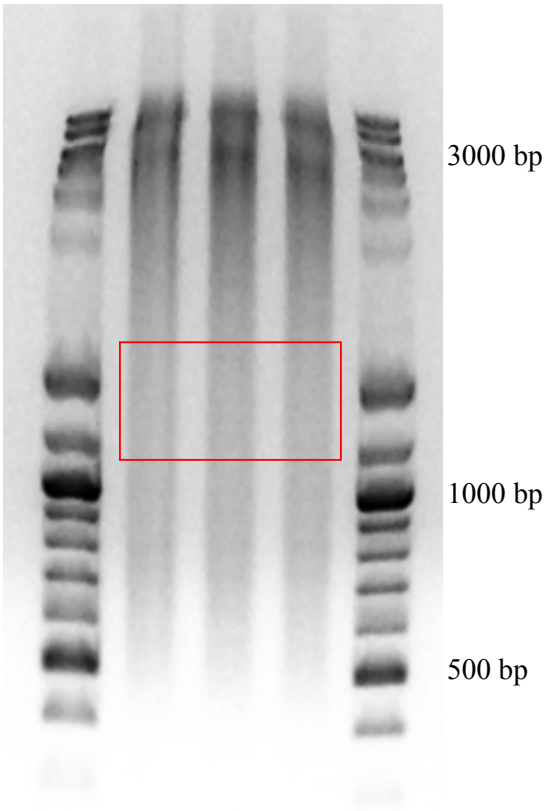


Figure 4. A 1% agarose gel showing the results of multimerization of the **elastin-CCC** monomer DNA. Lane 1: Marker (NEB, 1 kb, 100 bp); Lanes 2, 3, 4: Multimerization reaction; Lane 5: Marker (NEB, 1 kb, 100 bp). The red box is drawn around the region of the gel that was excised for isolation of multimers in the range of 2000-1200 bp in length.

The synthetic gene (**elastin-CCC**) was designed for the production of an elastin-mimetic polypeptide in which each proline residue in the elastomeric pentapeptide repeat sequence (VPGVG) was encoded by the CCC codon. The CCC codon is rarely used in the *E. coli* host organism, and the **elastin-CCC** gene was designed to take advantage of the inherent codon bias in the host organism and enable codon capture for incorporation of proline analogues in the elastin-mimetic polypeptides using an orthogonal system. The orthogonal system used for site-specific incorporation of proline analogues is described in detail in Chapters 3 and 4. Construction of the **elastin-CCC** gene was critical as a target for protein expression to evaluate the activity of the orthogonal system. As the **elastin-CCC** gene was based on the sequence of **elastin-1**[24], the **elastin-CCC** protein product is similar in sequence and structure to the previously studied **elastin-1** protein. Furthermore, the **elastin-CCC** gene will serve as a model system for biosynthesis of collagen-mimetic proteins. High level, accurate co-translational incorporation of proline analogues, such as *trans*-4-hydroxy-L-proline, in addition to incorporation of L-proline in the same polypeptide is one of the major future goals for the project. The **elastin-CCC** coding sequence is of particular importance in construction of the orthogonal system that would enable site-specific incorporation of hydroxyproline in elastin-mimetic materials.

The plasmid vectors generated in the synthesis of the **elastin-CCC** gene are shown in Figures 5, 6, 7 and 8. The high copy number pQE-80L plasmid was the vector of choice for high-level production of the **elastin-CCC** protein as the plasmid includes a copy of the overproducing repressor allele, $lacI_q$, to ensure tight control of the basal level of transcription prior to induction of **elastin-CCC** protein expression with IPTG.

Ampicillin resistance is enabled by the activity of a β-lactamase encoded on the plasmid that is used for plasmid maintenance and colony selection. The adaptor sequence was designed to include a decahistidine tag at the C-terminus of the **elastin-CCC** gene to facilitate protein purification using a metal affinity column as described in Chapter 3.

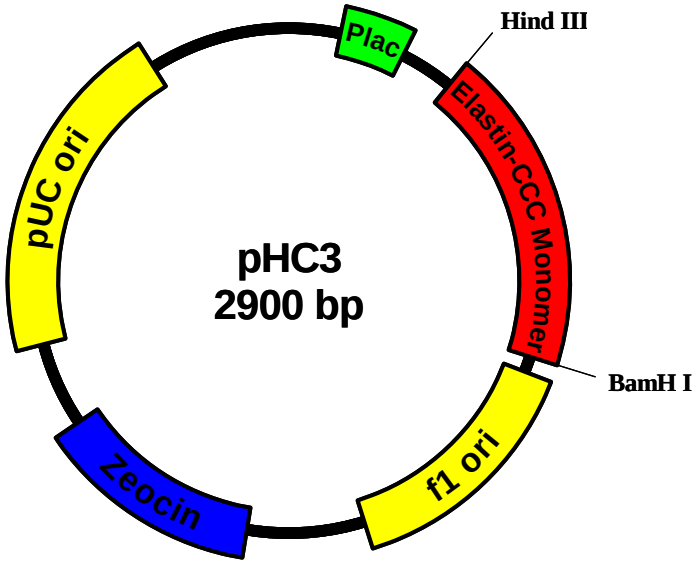


Figure 5. Plasmid pHC3 represents the pZErO®-1plasmid containing the synthetic **elastin-CCC** monomer gene sequence.

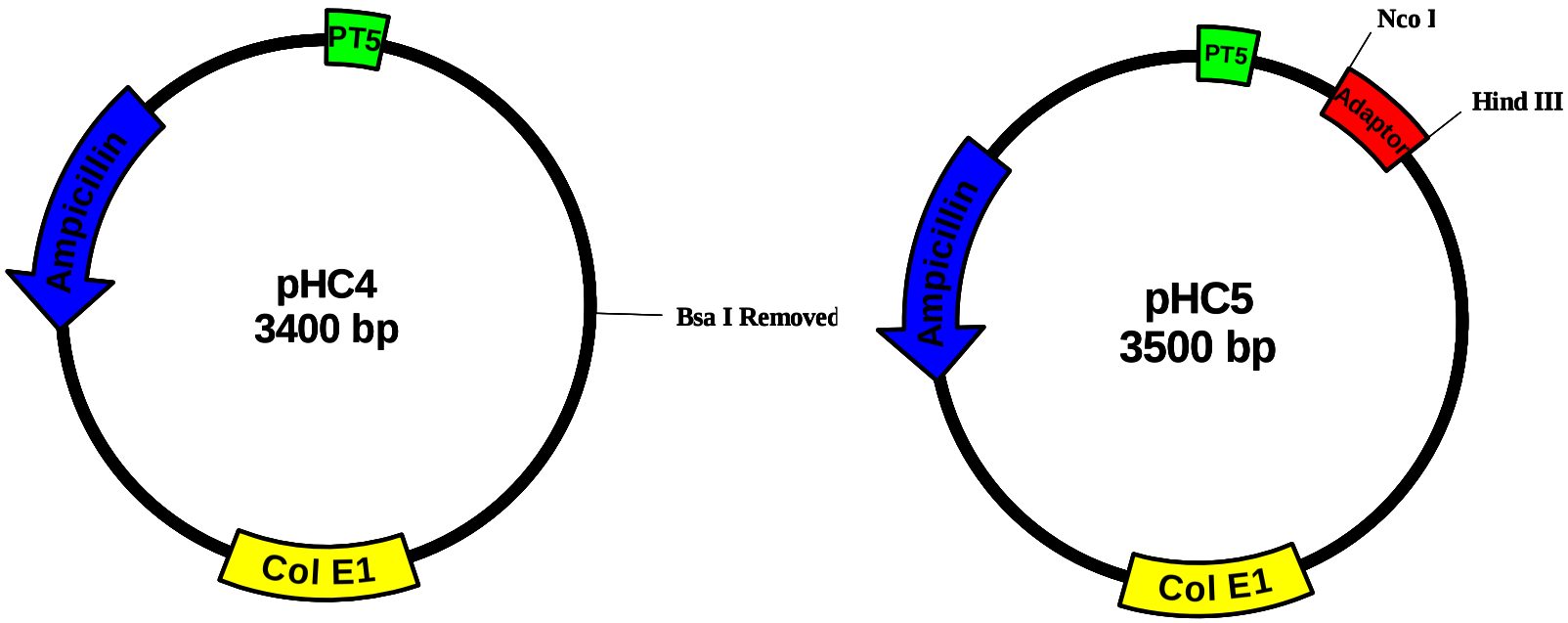


Figure 6. The pHC4 plasmid was generated by site-directed mutagenesis of the pQE-60 plasmid to remove a *Bsa* I restriction site to enable cloning into the expression plasmid. pHC5 was generated by cloning the adaptor sequence into plasmid pHC4 utilizing the restriction cleavage sites present in the multiple cloning site of plasmid pQE-60.

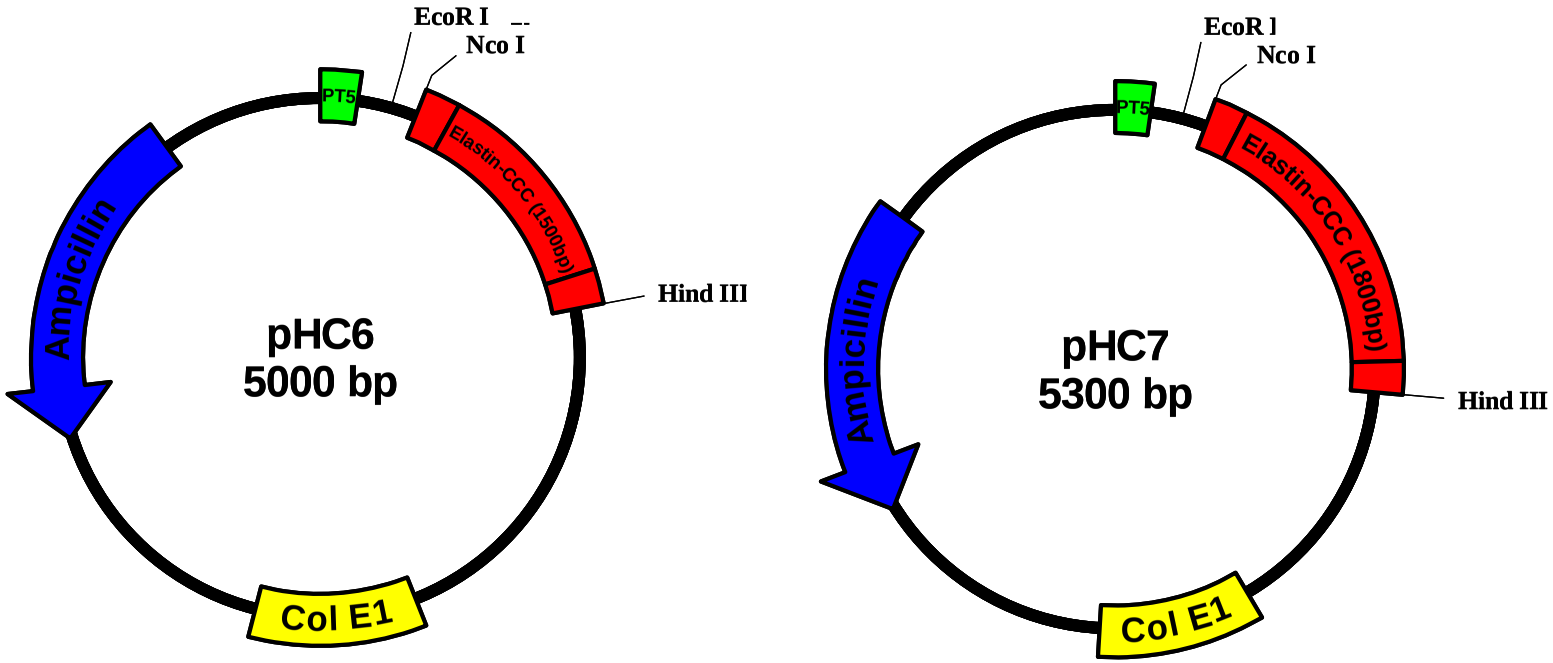


Figure 7. The pHC6 plasmid was generated by cloning the **elastin-CCC** multimer gene (length of ~1500 bp) into the pHC5 acceptor plasmid. The pHC7 plasmid was generated by cloning the **elastin-CCC** multimer gene (length of ~1800 bp) into the pHC5 acceptor plasmid.

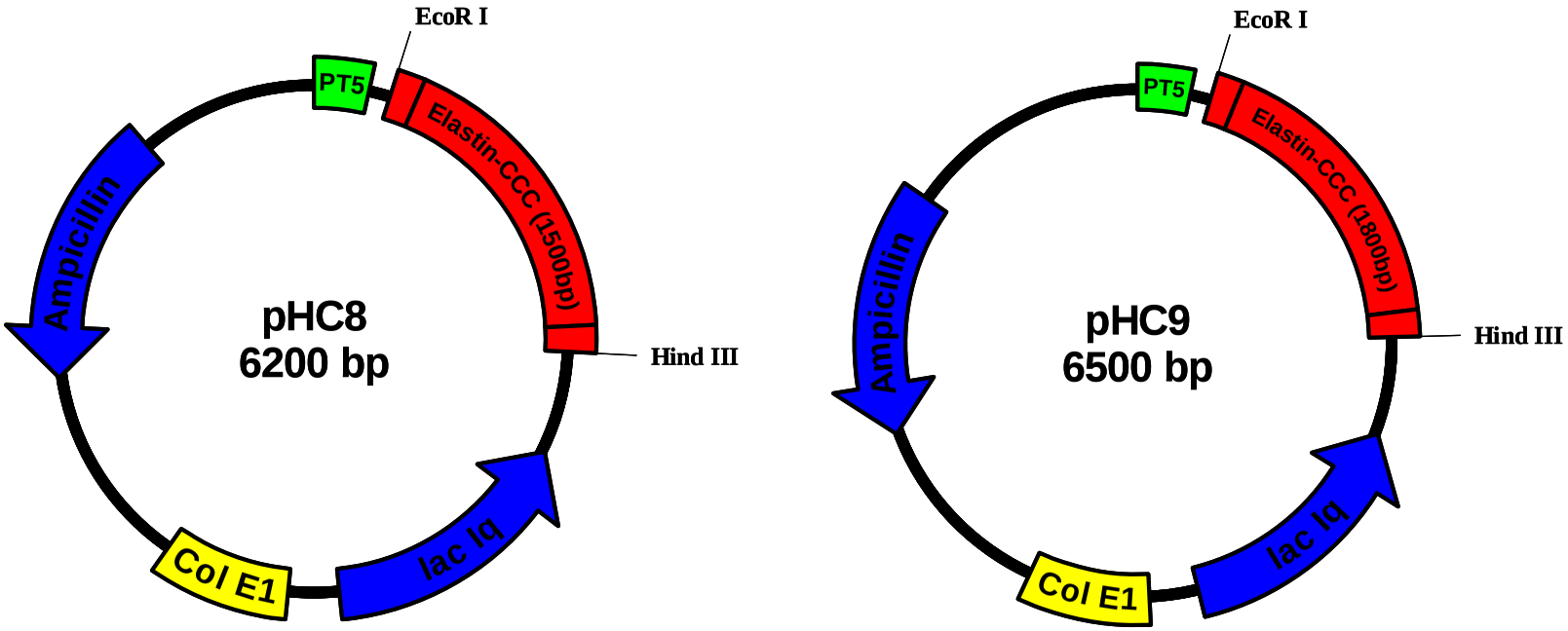


Figure 8. The pHC8 plasmid was generated by cloning the **elastin-CCC** multimer gene cassette (length of ~1500 bp) into the pQE-80L expression plasmid. The pHC9 plasmid was generated by cloning the **elastin-CCC** multimer gene cassette (length of ~1800 bp) into the pQE-80L expression plasmid.

Elastin-CCC: $M[(VPGVG)_4VPGIG]_{17}VPGVGGSD_4KGH_{10}$

Scheme 1. Elastin-CCC polypeptide sequence based on translation of the gene sequence.

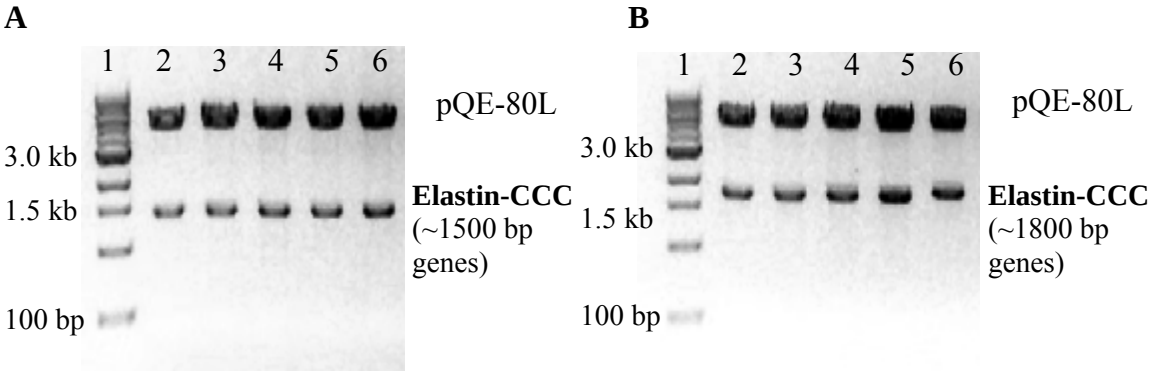


Figure 9. A. A 1 % agarose gel depicting plasmid pHC8 double digested with *Eco*R I/Hind III indicating the **elastin-CCC** gene insert. Five microliters of each digestion reaction were run on the gel in separate lanes. Lane 1: 1 kb, 100 bp DNA markers (NEB); Lanes 2-6: The DNA from 5 single colonies, all containing pHC8, was isolated for digestions. **B.** 1 % agarose gel depicting plasmid pHC9 double digested with *Eco*R I/*Hind* III indicating the **elastin-CCC** gene insert. Lane 1: 1 kb, 100 bp DNA markers (NEB); Lanes 2-6: The DNA from 5 single colonies, all containing pHC9, was isolated for digestions.

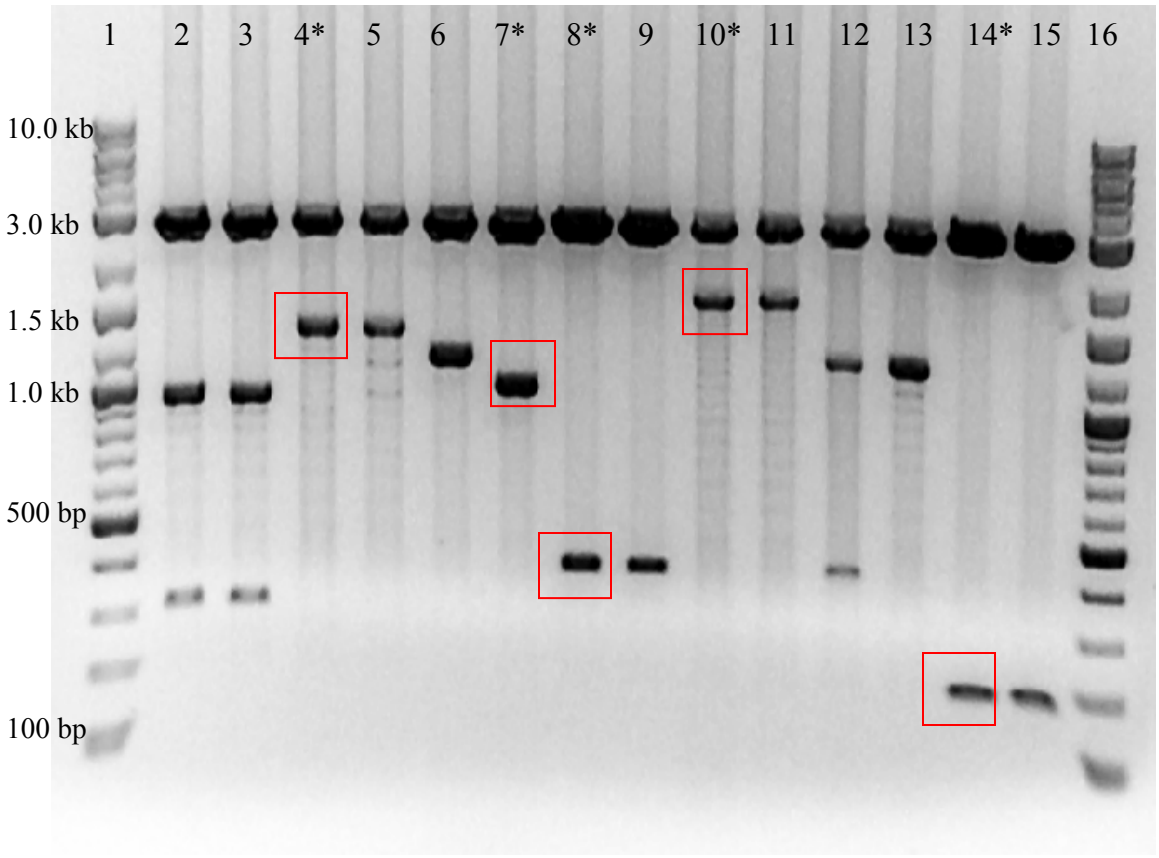


Figure 10. A 1 % agarose gel demonstrating the results of multimerization. DNA was isolated from several separate colonies grown on selective plates after transformation with plasmids containing **elastin-CCC** multimers of various lengths. The DNA was digested in 14 separate reactions and 5 µL of each reaction were run on the gel. Several multimers of variable lengths were chosen for cloning and future experiments. Lanes marked with an asterisk (*) show multimers that were isolated for future analysis. The boxed regions correspond to multimers of the following approximate sizes, 1400 bp, 1100 bp, 500 bp, 2000 bp, and 200 bp. A faint ladder can be seen in many of the lanes possible indicating that the gene is unstable in the pZErO®-1 plasmid. Lanes 1 and 16: 1 kb, 100 bp DNA markers (NEB); Lanes 2-15: The DNA from 14 single colonies, each containing the pZErO®-1 plasmid and **elastin-CCC** multimer insert.

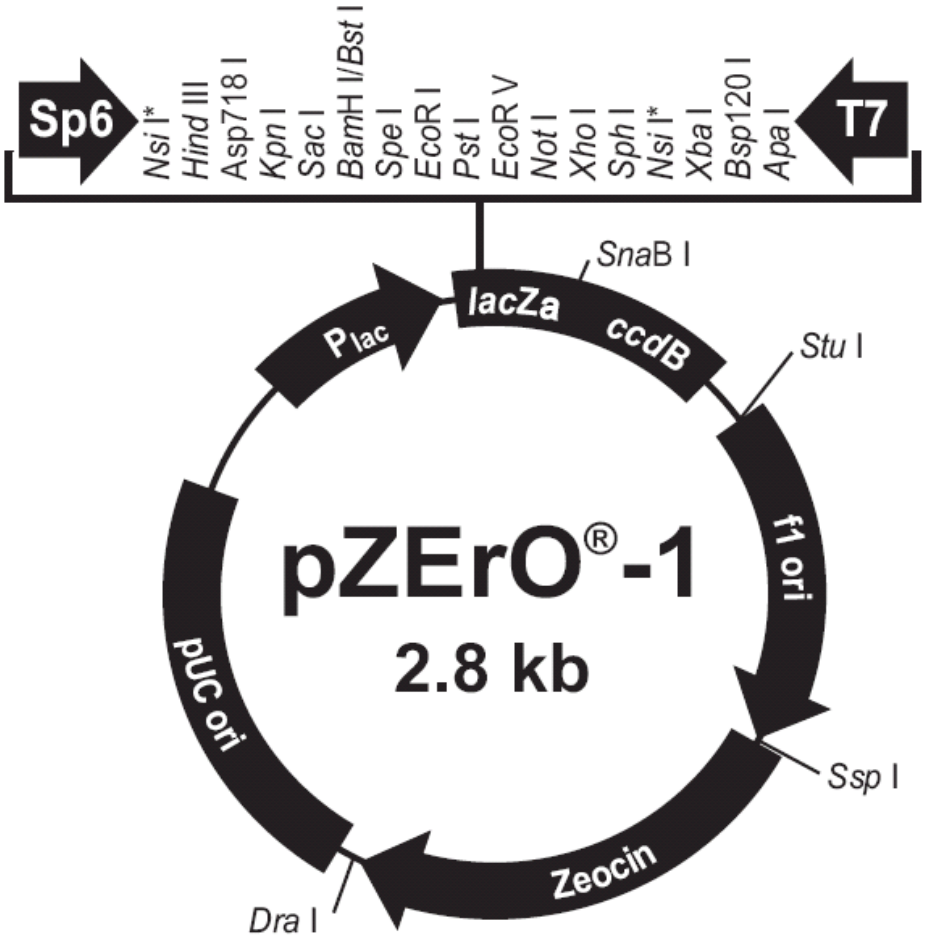


Figure 11. Plasmid map of the pZErO®-1 vector, from Invitrogen, Inc., indicating restriction endonuclease cleavage sites within the multiple cloning site. The **elastin-CCC** adaptor DNA sequence was cloned into the plasmid using the *Bam*H I/*Hind* III sites.

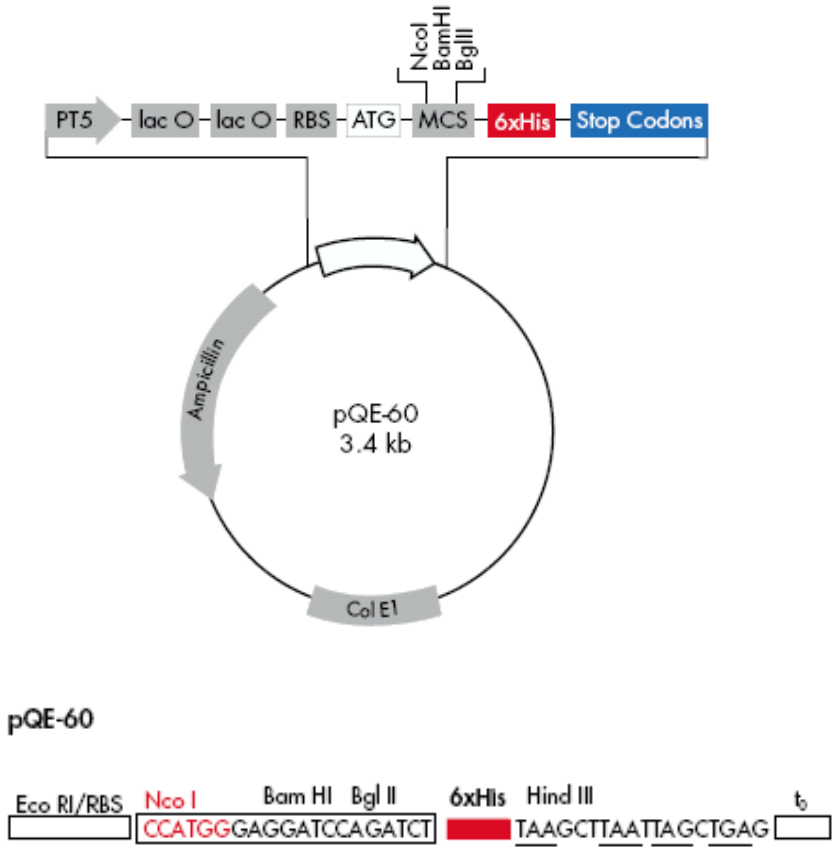


Figure 12. Above Plasmid map of the pQE60 vector, from QIAGEN, indicating restriction endonuclease cleavage sites within the multiple cloning site. **Below** Expanded detail of the multiple cloning site (MCS).

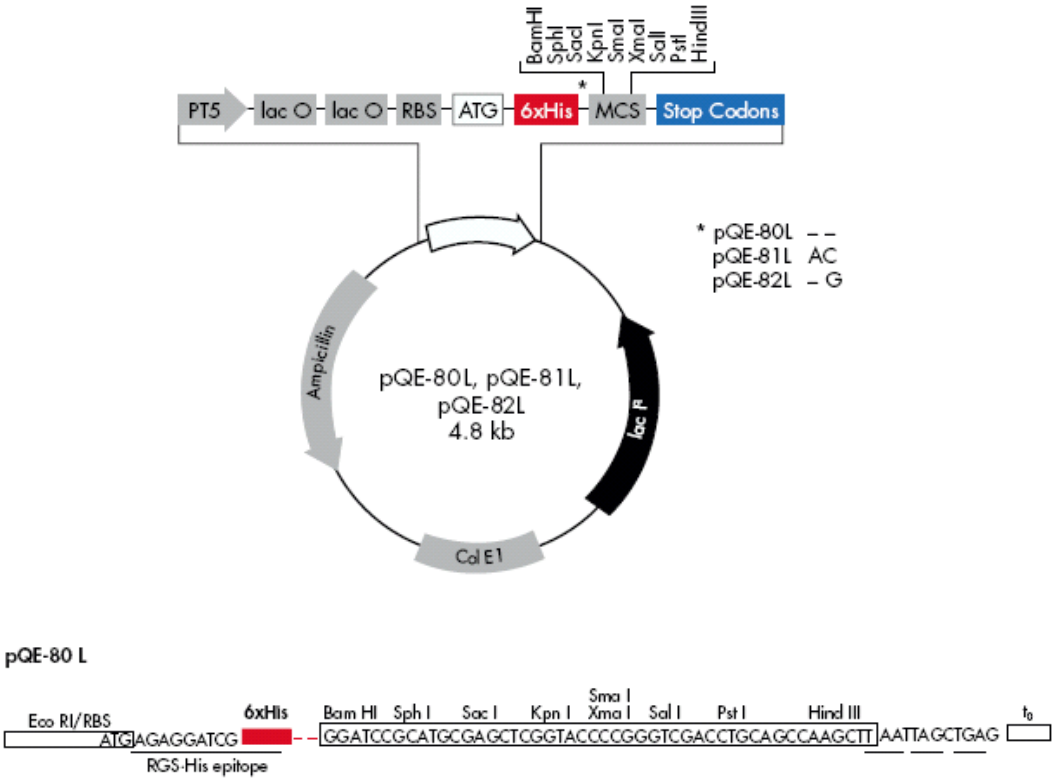


Figure 13. Above Plasmid map of the pQE-80L vector, from QIAGEN, indicating restriction endonuclease cleavage sites within the multiple cloning site. **Below** Expanded detail of the multiple cloning site (MCS).

Conclusions

Genetic engineering techniques were useful in the design and engineering of an elastin-mimetic protein polymer with defined sequence specificity. By using these methods, one can control at the DNA level, the amino acid sequence of polypeptide materials for high-level, cost-effective biosynthesis in a bacterial host. Control over sequence specificity is essential for further applications including utilizing the inherent

codon bias in the *E. coli* host for site-specific incorporation of amino acid analogues at defined positions throughout the elastin-mimetic product of interest.

References

(1) van Hest, J. C. M.; Tirrell, D. A. *Chemical Communications* **2001**, 1897-1904.

(2) Langer, R.; Tirrell, D. A. *Nature* **2004**, *428*, 487-492.

(3) Maskarinec, S. A.; Tirrell, D. A. *Current Opinion in Biotechnology* **2005**, *16*, 422-426.

(4) Nagapudi, K.; Brinkman, W. T.; Thomas, B. S.; Wright, E. R.; Conticello, V. P.; Chaikof, E. L. *Cell Transplantation* **2003**, *12*, 170-171.

(5) Kim, W.; Conticello, V. P. *Polymer Reviews* **2007**, *47*, 93-119.

(6) McMillan, R. A.; Lee, T. A. T.; Conticello, V. P. *Macromolecules* **1999**, *32*, 3643-3648.

(7) Huang, L.; McMillan, R. A.; Apkarian, R. P.; Pourdeyhimi, B.; Conticello, V. P.; Chaikof, E. L. *Macromolecules* **2000**, *33*, 2989-2997.

(8) McMillan, R. A.; Conticello, V. P. *Macromolecules* **2000**, *33*, 4809-4821.

(9) Qu, Y.; Payne, S. C.; Apkarian, R. P.; Conticello, V. P. *Journal of the American Chemical Society* **2000**, *122*, 5014-5015.

(10) Zhou, Y. T.; Wu, S. X.; Conticello, V. P. *Biomacromolecules* **2001**, *2*, 111-125.

(11) Wright, E. R.; Conticello, V. P. *Advanced Drug Delivery Reviews* **2002**, *54*, 1057-1073.

(12) Wright, E. R.; McMillan, R. A.; Cooper, A.; Apkarian, R. P.; Conticello, V. P. *Advanced Functional Materials* **2002**, *12*, 149-154.

(13) Nagapudi, K.; Brinkman, W. T.; Leisen, J.; Thomas, B. S.; Wright, E. R.; Haller, C.; Wu, X. Y.; Apkarian, R. P.; Conticello, V. P.; Chaikof, E. L. *Macromolecules* **2005**, *38*, 345-354.

(14) Nagapudi, K.; Brinkman, W. T.; Thomas, B. S.; Park, J. O.; Srinivasarao, M.; Wright, E.; Conticello, V. P.; Chaikof, E. L. *Biomaterials* **2005**, *26*, 4695-4706.

(15) Wright, E. R.; Conticello, V. P.; Apkarian, R. P. *Microscopy and Microanalysis* **2003**, *9*, 376-376.

(16) Sallach, R. E.; Jordan, S. W.; Apkarian, R. P.; Conticello, V. P.; Chaikof, E. L. *Scanning* **2006**, *28*, 114-115.

(17) Sallach, R. E.; Wei, M.; Biswas, N.; Conticello, V. P.; Lecommandoux, S.; Dluhy, R. A.; Chaikof, E. L. *Journal of the American Chemical Society* **2006**, *128*, 12014-12019.

(18) Tirrell, D. A.; Fournier, M. J.; Mason, T. L. *MRS Bulletin* **1991**, *16*, 23-28.

(19) Goeden-Wood, N. L.; Conticello, V. P.; Muller, S. J.; Keasling, J. D. *Biomacromolecules* **2002**, *3*, 874-879.

(20) McMillan, R. A.; Caran, K. L.; Apkarian, R. P.; Conticello, V. P. *Macromolecules* **1999**, *32*, 9067-9070.

(21) Wright, E. R.; Conticello, V. P.; Apkarian, R. P. *Microscopy and Microanalysis* **2003**, *9*, 171-182.

(22) Urry, D. W. *Journal of Physical Chemistry B* **1997**, *101*, 11007-11028.

(23) Sambrook, J.; Russell, D. W. *Molecular Cloning: A laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

(24) Kim, W. Y.; George, A.; Evans, M.; Conticello, V. P. *Chembiochem* **2004**, *5*, 928-936.

Chapter 3

Design and Genetic Engineering of a Novel Orthogonal Translational System in *E. coli*

Introduction

The expansion of the genetic code of *E. coli* has recently been achieved using directed evolution techniques in a variety of applications[1]. Amino acid analogues have been incorporated directly into proteins in bacterial, fungal, and mammalian cell hosts[1-6]. Direct, co-translational incorporation of amino acid analogues gives an advantage over post-translational modification of polypeptide sequences in that one is able to achieve high protein expression levels in a host organism with highly homogenous molecular sequences. An *in vivo* system for production of elastin-mimetic and collagen-mimetic polypeptides would enable co-translational, site-specific incorporation of proline or proline analogues, such as hydroxyproline, in response to a defined DNA coding sequence. Hydroxyproline residues, abundantly present in collagen, are normally formed through post-translational hydroxylation of proline by the enzyme, prolyl-4-hydroxylase. Post-translational modification is limited in the inefficient or inconsistent chemical modification of proline residues leading to products of variable sequence and in resultant low protein yields[7]. *In vitro* incorporation of various other analogues has been implemented[8-10], but using a host organism would allow for more efficient and highly accurate production of biomaterials through biosynthesis[11,12].

An aminoacyl-tRNA synthetase/tRNA orthogonal pair can be used to independently incorporate amino acid analogues in a host organism in response to a specific DNA coding sequence (Scheme 1). The pair is chosen from an organism with relatively little similarity to the target host organism. For example, an archaeal, *M. jannaschii* tRNA/synthetase pair, compared to an *E. coli* host, exhibits decreased sequence similarity to the host machinery, and one would expect to see different

aminoacyl-tRNA synthetase recognition elements between the two species[13,14]. An orthogonal tRNA/synthetase pair is introduced into the host using genetic engineering techniques, and the pair functions independently in the host organism as long as cross-species aminoacylation is inefficient and the anticodon loop is the only significant determinant of synthetase/tRNA recognition. One such source for orthogonal aminoacyl-tRNA synthetase/tRNA pairs is the archaeal *Methanococcus jannaschii* bacterial species. *M. jannaschii* has been used successfully as a source for orthogonal pairs in several incorporation studies[4,15-18]. A human orthogonal pair expressed in *E. coli* would likely be ideal for orthogonality[19] and potentially for production of human genes such as elastin and collagen, but a human synthetase and tRNA may exhibit limitations in expression and activity in a bacterial expression host.

Methods to expand the genetic code have included the selection and evolution of orthogonal aminoacyl-tRNA synthetase/tRNA (tRNA/aaRS) pairs that function in a heterologous host, such that a non-canonical amino acid is inserted in response to a termination codon using non-sense suppression[20]. Many previous investigations into the incorporation of amino acid analogues have been focused on stop codon, particularly amber codon suppression in addition to four-base codon supression[20-23]. Peter Schultz, in particular, has made advances in the use of orthogonal tRNA/synthetase pairs for incorporation of a variety of analogues using the amber codon suppression system[1]. While successful incorporation of non-canonical amino acids has been achieved, non-sense suppression has limited efficiency, particularly for multi-site incorporation of amino acid analogues[11,24]. Therefore, a different approach was designed for high-level incorporation of amino acid analogues in protein biomaterials.

Codon usage statistics in *E. coli* suggest that a subset of the canonical codons occur very rarely in structural genes within the genome[25]. These codons are good candidates for capture using an orthogonal tRNA/aaRS system, as their use should not induce termination or significantly disrupt the normal metabolism of the host organism[24]. The CCC codon of the proline family box was chosen as a target site for incorporation of non-canonical proline analogues in an *E. coli* expression system. This approach was based upon the observations that mutant strains of the related organism *S. typhimurium* containing a dysfunctional version[26] or deletion[27] of the gene encoding the native tRNAPro(GGG) are viable. Furthermore, the CCC codon occurs at a very low frequency within structural genes in the *E. coli* genome (Table 1), selective methods have been developed for knocking out specific tRNA genes in the *E. coli* genome[28], and orthogonal tRNAPro/ProRS pairs exist in *H. sapiens*[19] and *M. jannaschii*[13]. An orthogonal pair would enable co-translational incorporation of proline analogues, such as (2*S*,4*R*)-4-hydroxyproline or (2*S*,4*R*)-4-fluoroproline, in response to the CCC codon while canonical proline incorporation would be achieved in response to the other codons of the proline family box[24] (Scheme 1). Chapter 4 gives a detailed discussion of development of a mutant version of a prolyl-tRNA synthetase that is selective for specific proline analogues in comparison to the canonical amino acid. To date, there have been no published studies of orthogonal systems developed for the site-specific incorporation of proline or proline analogues.

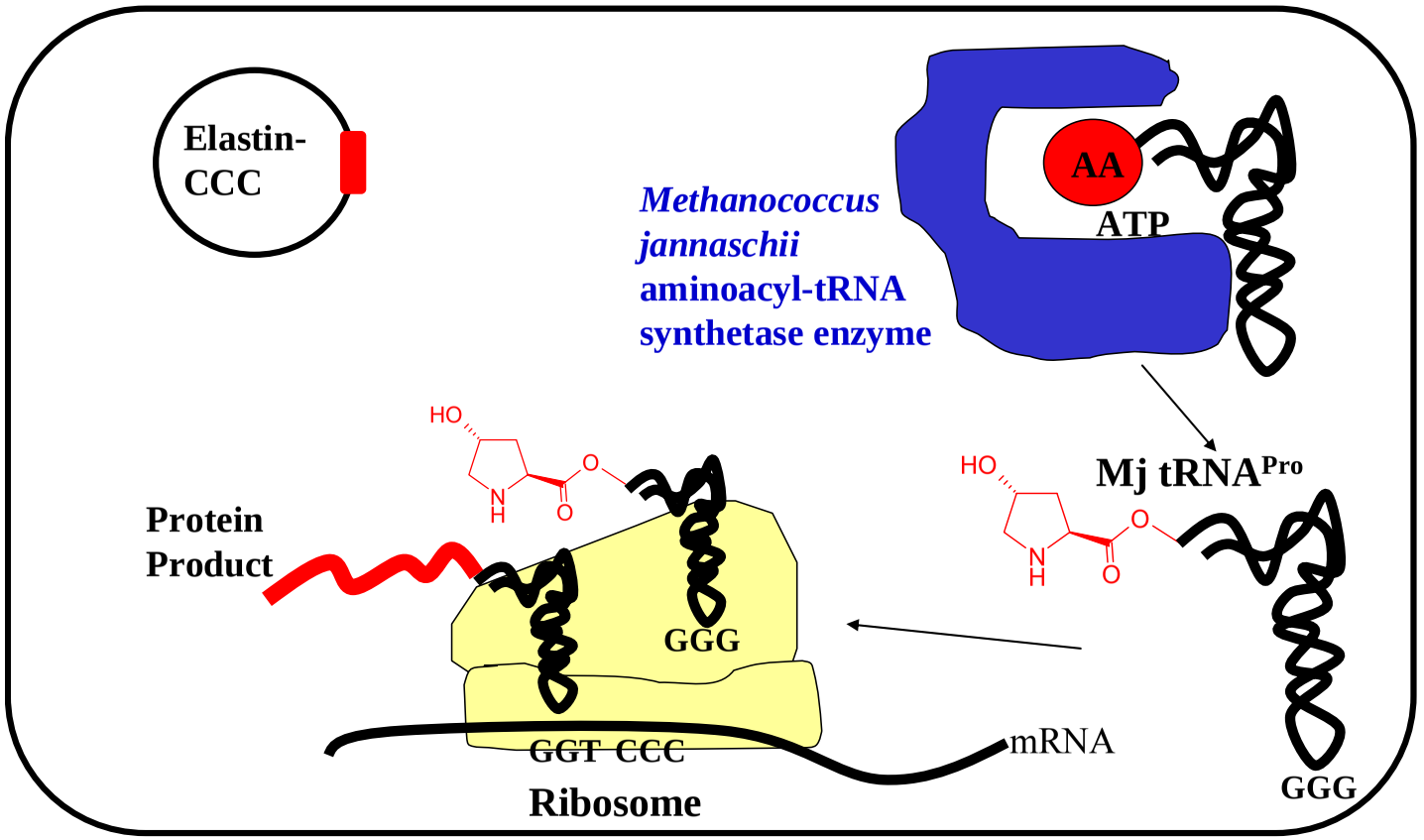


Figure 1. Using an orthogonal system for incorporation of (*2S,4R*)-4-hydroxyproline in an *E. coli* bacterial host. The **elastin-CCC** gene is encoded by an expression vector present in the host cell. The aminoacyl-tRNA synthetase/tRNA pair is encoded on a separate expression plasmid to enable incorporation of hydroxyproline in **elastin-CCC** in response to the CCC codon upon induction. Other amino acids are incorporated using the endogenous cellular machinery in the host.

Multi-site, global substitution with a variety of proline analogues for native proline was achieved in elastin-mimetic, proline-rich polypeptide sequences with protein expression yields ranging from 15-50 mg/L[29]. In contrast, site-specific substitution rather than global substitution is essential for the production of native collagen and/or synthetic collagen analogues as both proline and hydroxyproline must be selectively incorporated. The majority of collagen used in biomedical applications is currently derived from bovine or porcine sources[30]. Generating polypeptide analogues, which retain the useful properties of native collagen through a biosynthetic approach such as protein engineering, provides sequence control, minimizes batch-to-batch variation, and avoids contamination from pathogens. Biosynthetic methods potentially provide freedom from the necessity for post-translational modification of the synthetic protein *in vivo*.

Development of the orthogonal system, using an elastin-mimetic protein, as a model for collagen, is desirable due to the extensive experience our lab has with design and production of elastin-mimetic polypeptides. Elastin-mimetics are characterized by a high proline content and repetitive sequence, as in collagen, and they have been expressed in a bacterial host. Elastin-mimetic proteins have the advantage that the synthetic genes can be readily maintained and expressed to high yield in bacterial host systems. In addition, incorporation of hydroxyproline and other proline analogues into elastin-mimetic proteins has been demonstrated in *E. coli* with protein yields on the order of 15-50 mg protein/L of culture[29]. Proline and proline analogues, such as hydroxyproline, are of particular interest for design of a system for the biosynthesis of elastin-mimetic and collagen-mimetic polypeptides where incorporation of

hydroxyproline or other analogues can be controlled at the DNA level using a novel orthogonal translational system.

Codon	*E. coli*	*H. sapiens*	*S. cerevisiae*	*M. jannaschii*	*T. thermophilus*
CCU	7.1	17.4	13.5	6.9	3.9
CCC	5.5	20.0	6.8	1.4	45.6
CCA	8.5	16.8	18.2	31.9	1.4
CCG	23.2	7.0	5.3	1.4	11.7

Table 1. Codon usage within the proline family box (codon usage in frequency per thousand). The CCC codon is rarely used for encoding proline within *E. coli* structural genes.

Materials

All chemical reagents were purchased from either Fisher Scientific, Inc. (Pittsburgh, PA) or Sigma Chemical Co. (St. Louis, MO) unless otherwise noted. Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases, T4 DNA ligase, and T4 kinase were purchased from New England Biolabs, Inc. (Beverly, MA), shrimp alkaline phosphatase was obtained from Roche Applied Science (Indianapolis, IN), and Platinum *Pfx* DNA polymerase was obtained from Invitrogen Corp. (Carlsbad, CA). Plasmid pPROTetE.133 was obtained from BD Biosciences, Inc. (Palo Alto, CA). Plasmid pPROLarA.231 was obtained from Professor Rik Myers of the University of Miami. Plasmid pCS-364, which encodes the native *E. coli* prolyl-tRNA synthetase as an *N*-terminal hexahistidine fusion in plasmid pQE-30 (QIAGEN, Inc.), was a gift from Professor Karin-Musier Forsyth of the University of Minnesota. *E. coli* strain DG99 was purchased from American Type Culture Collection (ATCC# 47041) and strains CAG18515[31] and UQ27[32] were obtained from the *E. coli* Genetic Stock Center at Yale University. The *E. coli* strain, Top10F´ was obtained from Invitrogen Corp. (Carlsbad, CA). The pKS509 plasmid was a gift from Professor Karin-Musier Forsyth of the University of Minnesota[19]. The pSU81 plasmid was a gift from Professor William McClain at the University of Wisconsin[33]. The plasmid pRAM2 was employed as a source of the gene encoding the **elastin-1** sequence. Double digestion of pRAM2 with *Nco* I and *Bam*H I afforded a duplex DNA cassette of approximately 1300 bp, which was inserted into the compatible *Nco* I/*Bgl* II sites of plasmid pQE-60 (QIAGEN, Inc., Valencia, CA) to generate plasmid pAG1. An *Eco*R I/*Hin*D III cassette derived from

pAG1 was excised, isolated, and cloned into the compatible sites of plasmid pQE-80L (QIAGEN, Inc., Valencia, CA) to afford pAG2. Plasmid constructs encoding variants of the *E. coli* prolyl-tRNA synthetase were derived from plasmid pCS-364 (a gift from Professor Karin-Musier Forsyth of the University of Minnesota), and plasmid pWK1 encoding the EcProRS gene was generated by Dr. Wookhyun Kim in the Conticello lab at Emory University[29].

Synthetic oligonucleotides were purchased from either Sigma-Genosys, Inc (The Woodlands, TX) or Integrated DNA Technologies (Coralville, IA) and were used as received. TALON® metal affinity resin was purchased from Clontech, Inc. (Mountain View, CA). NMM medium was prepared as described previously[29], with the exception that proline was not added to the medium prior to cell culture. Similarly, SMM media was prepared according to the protocol by Studier[34].

Procedures for the manipulation of DNA, the transformation of competent cells, and the growth and induction of bacterial cultures were adapted from the published literature[35] or instructions supplied by manufacturers. Reagents for the manipulation of bacteria and DNA were sterilized by either autoclave or passage through a 0.2 μm filter. Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. Site-directed mutagenesis was performed using Stratagene's (La Jolla, CA) Quick-Change mutagenesis technique from gene-specific oligonucleotide primers. Automated DNA sequencing was performed at the Emory University Core DNA Sequencing Facility on a Perkin-Elmer ABI Prism model 377 DNA sequencer.

Protein electrophoresis was performed using 10-15 % gradient discontinuous pre-cast sodium dodecyl sulfate (SDS) polyacrylamide gels on a PhastSystem from GE

Healthcare Bio-Sciences Corp. (Piscataway, NJ). The Perfect Protein Marker was used as a protein standard for SDS-PAGE and western blot analysis and was purchased from Novagen, Inc. (Madison, WI). Silver staining, used for protein visualization, was carried out using a silver staining kit purchased from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ). Western blotting was carried out using the PhastSystem (GE Healthcare). The primary antibody used for binding his-tagged proteins was the His-tag monoclonal antibody from Clontech, Inc. (Mountain View, CA). The secondary antibody used was the goat anti-mouse secondary antibody, also from Clontech, Inc. Polyhistidine tagged proteins were visualized by chemiluminescent detection using the 1-step NBT/BCIP reagent mixture from Pierce Biotechnology (Thermo Scientific) (Rockford, IL).

Methods

Preparation of the Bacterial Host Strain

A P1 transduction was performed to confer a proline auxotrophic phenotype to the UQ27 strain. A 20 µL aliquot of an overnight culture of ΔPro bacterial cells was added to 2 mL of LB media supplemented with 20 µL 20% glucose and 10 µL $CaCl_2$. The cell suspension was incubated on a rotator for 30 min at 37 °C. A 20 µL portion of a P1 *dam rev*6 phage suspension (~10^9/mL) was placed at 37 ºC for 3 min in an open-top tube to remove $CHCl_3$. The phage suspension was added to the cell suspension and incubated on a rotator for 2-3 h until lysis had occurred (suspension was clear). Three drops $CHCl_3$ were added to the cells and the tube was vortexed gently. The suspension was centrifuged at 10,000 xg for 5 min, and the supernatant was stored over $CHCl_3$ at 4 ºC.

The tube was inverted several times to kill any residual bacteria. Ten µL of the donor phage suspension were then placed at 37 ºC for 3 min in an open-top tube to remove $CHCl_3$. One mL of SOC rich media, supplemented with 5 mM $CaCl_2$ and 10 mM $MgSO_4$, was added to the phage suspension. A control reaction was prepared in the same way, but no donor phage was added to the tube. A 200 µL aliquot of an overnight culture of UQ27 was added to each of the two 13 x 100 mm screw cap reaction tubes. The suspensions were mixed by inversion and incubated for 6 h at 30 °C. The cells were centrifuged at 4000 x g for 10 min at room temperature, and the supernatant was discarded. Cells were resuspended in 2 mL SOC broth, and 150 µL were spread onto solid LB media supplemented with kanamycin (10 µg/mL). No colonies were observed on the negative control plate.

After 24 h, many colonies were seen on the experimental plate. Several colonies from the plate were re-streaked on solid LB media (with kanamycin at a concentration of 10 µg/mL and 50 µg/mL to evaluate antibiotic resistance. Cells that grew on the selective plates were resistant at both concentrations due to uptake of the kanamycin resistance gene with concomitant proline auxotrophy. Single colonies were picked from the selective plates and were grown in new minimal media either in the presence of proline or without proline. After 16 h growth, shaking at 225 rpm, at 30 ºC, no growth was seen in the media lacking proline for each of the selected colonies.

Plasmid and Gene Construction

***Homo sapiens* prolyl-tRNA synthetase/tRNA (ProRS/$tRNA^{Pro}$) pair**

An acceptor expression plasmid (plasmid pPROTet/Lar) for the prolyl-tRNA synthetase (ProRS) and tRNA genes was constructed from ligation of the *Avr* II/*Spe* I

fragment of pPROTetE.133, containing the transcriptional/translational control elements, multiple cloning site, and chloramphenicol resistance gene, to the corresponding fragment of pPROLarA.231, containing the p15A origin of replication (producing pME1). It was first necessary to introduce *Nhe* I and *Pvu* I restriction endonuclease sites in the plasmid, pPROTet/Lar (pME1), for cloning of the tRNA gene cassettes. Inverse PCR amplification was used to introduce the sites into the plasmid using the Lar-F and Lar-R primers with the sequences shown below:

Lar-F: 5′-CGAATCCCGGGCGATCGCGGAGATTTCCTGGAAGATGCCAGG-3′

Lar-R: 5′-AGCTACCCGGGCTAGCCCGTTCGTAAGCCATTTCCGCTCGC-3′

These primers were designed to introduce a *Xma* I cleavage site between the *Nhe* I and *Pvu* I sites in the plasmid for linearization and subsequent re-ligation of the plasmid after inverse PCR amplification of the entire plasmid. The final product, pHEC1, contained the *Nhe* I and *Pvu* I sites in a non-coding region of the plasmid for cloning of the tRNA gene cassettes.

A recombinant copy of the human ProRS was amplified from the plasmid, pKS509, using gene specific primers that were designed to introduce *Kpn* I and *Xba* I restriction sites at the 5′ and 3′ termini of the HsProRS gene. After amplification, the gene product was double-digested with *Kpn* I/*Xba* I and the DNA was purified by gel isolation. The HsProRS was then ligated into the *Kpn* I/*Xba* I double-digested acceptor plasmid, pHEC1. The sequence of the HsProRS sequence, in the new plasmid, pHC1, was confirmed by automated DNA sequencing analysis using primers designed to bind to the pHEC1 plasmid upstream and downstream of the multiple cloning site. The human tRNAPro was constructed by annealing oligonucleotide primers containing the entire

HstRNA^{Pro} gene sequence. The primers were designed to incorporate *Eco*R I and *Pst* I sticky ends, 4 nucleotide base overhangs, for direct cloning of the tRNA gene into the acceptor plasmid, pSU81.

tRNAPro-F

5′-AATTCGGCTCGTTGGTCTAGGGGTATGATTCTCGCTTGGGGTGCGAG

AGGTCCCGGGTTCAAATCCCGGACGAGCCCCCACTG-3′

tRNAPro-R

5′-GTGGGGGCTCGTCCGGGATTTGAACCCGGGACCTCTCGCACCCCAA

GCGAGAATCATACCCCTAGACCAACGAGC-3′

The forward and reverse primers corresponding to the HstRNA^{Pro} sequence were annealed using a BioRad MJ Mini thermocycler (Hercules, CA). Annealing was carried out by dissolving the primers in distilled, deionized water (ddH_2O) to a final concentration of 0.5 μg/μL. Ten μL of each of the two primers were mixed together with 4 μL 5 M NaCl, 4 μL $MgCl_2$, and 172 μL ddH_2O. By gradually decreasing the temperature of the reactions from 99 °C to 30 ºC (decreased by 5 °C every 3 min), the DNA strands were annealed together. The double-stranded DNA fragment was visualized by DNA agarose gel electrophoresis (4 % NuSieve agarose). The tRNA gene DNA was phosphorylated at the 5′ and 3′ positions using a T4 polynucleotide kinase enzyme. The pSU81 plasmid was double-digested with *Eco*R I/*Pst* I restriction endonuclease enzymes and the HstRNA gene DNA was ligated into the pSU81 plasmid. The sequence of the HstRNA sequence was confirmed by automated DNA sequencing analysis using primers designed to bind to the pSU81 plasmid upstream and downstream of the tRNA cloning sites.

The pSU81 plasmid, containing the HstRNA^{Pro} flanked by the *lpp* tRNA promoter and *rrnC* tRNA terminator sequences, was double-digested with *Nhe* I/*Pvu* I restriction endonucleases to isolate the tRNA expression cassette. The pHC1 plasmid, already containing the HsProRS gene, was also double-digested with *Nhe* I/*Pvu* I, and the tRNA cassette was ligated into the plasmid to generate plasmid pHC2. The sequences of the HsProRS and HstRNA^{Pro} genes were confirmed by automated DNA sequencing analysis.

***Methanococcus jannaschii* prolyl-tRNA synthetase/tRNA (ProRS/tRNA^{Pro}) pair**

The *M. jannaschii* ProRS was amplified from the *M. jannaschii* genome using gene specific primers that were designed to introduce *Kpn* I and *Xba* I restriction sites at the 5′ and 3′ termini of the MjProRS gene. After amplification, the gene product was double-digested with *Kpn* I/*Xba* I and the DNA was purified by gel isolation. The MjProRS was then ligated into the *Kpn* I/*Xba* I double-digested acceptor plasmid, pHEC1. The sequence of the MjProRS sequence, in the new plasmid, pHC10, was confirmed by automated DNA sequencing analysis using primers designed to bind to the pHEC1 plasmid upstream and downstream of the multiple cloning site. The MjProRS gene in plasmid, pHC10 and in the final expression plasmid, pHC15, is under the control of the constitutive P_L *tet* promoter. The archaeal tRNA^{Pro} was constructed by annealing oligonucleotide primers containing the entire MjtRNA^{Pro} gene sequence. The primers were designed to incorporate *Eco*R I and *Pst* I sticky end, 4 nucleotide base overhangs for direct cloning of the tRNA gene into the acceptor plasmid, pSU81.

MjtRNAPro-F

5′ - AATTCGGGGCCGTGGGGTAGCCTGGATATCCYGTGCGCTTGGGGGGC

GTGCGACCCGGGTTCAAGTCCCGGCGGCCCCACCACTGCA-3′

MjtRNAPro-R

5′- GTGGTGGGGCCGCCGGGACTTGAACCCGGGTCGCACGCCCCCCAAG

CGCACRGGATATCCAGGCTACCCCACGGCCCCG-3′

The forward and reverse primers corresponding to the $MjtRNA^{Pro}$ sequence were annealed using a BioRad MJ Mini thermocycler (Hercules, CA). Annealing was carried out by dissolving the primers in distilled, deionized water (ddH_2O) to a final concentration of 0.5 μg/μL. Ten μL of each of the two primers were mixed together with 4 μL 5 M NaCl, 4 μL $MgCl_2$, and 172 μL ddH_2O. By gradually decreasing the temperature of the reactions from 99 °C to 30 ºC (decreased by 5 °C every 3 min), the DNA strands were annealed together. The double-stranded DNA fragment was visualized by DNA agarose gel electrophoresis (4 % NuSieve agarose). The tRNA gene DNA was phosphorylated at the 5′ and 3′ positions using a T4 polynucleotide kinase enzyme. The pSU81 plasmid was double-digested with *Eco*R I/*Pst* I restriction endonuclease enzymes and the MjtRNA gene DNA was ligated into the pSU81 plasmid to generate plasmid, pHC13. The sequence of the MjtRNA sequence was confirmed by automated DNA sequencing analysis using primers designed to bind to the pSU81 plasmid upstream and downstream of the tRNA cloning sites.

The pSU81 plasmid, containing the $MjtRNA^{Pro}$ flanked by the *lpp* tRNA promoter and *rrnC* tRNA terminator sequences, was double-digested with *Nhe* I/*Pvu* I restriction endonucleases to isolate the tRNA expression cassette. The pHC10 plasmid, already containing the MjProRS gene, was also double-digested with *Nhe* I/*Pvu* I, and the tRNA cassette was ligated into the plasmid to generate plasmid pHC15. The sequences of the MjProRS and $MjtRNA^{Pro}$ genes were confirmed by automated DNA sequencing analysis.

Escherichia coli prolyl-tRNA synthetase/tRNA (ProRS/tRNA^{Pro}) pair

The *E. coli* ProRS was amplified from the pWK1 plasmid using gene specific primers that were designed to introduce *Kpn* I and *Xba* I restriction sites at the 5′ and 3′ termini of the EcProRS gene. After amplification, the gene product was double-digested with *Kpn* I/*Xba* I and the DNA was purified by gel isolation. The EcProRS was then ligated into the *Kpn* I/*Xba* I double-digested acceptor plasmid, pHEC1. The sequence of the EcProRS sequence, in the new plasmid, pHC19, was confirmed by automated DNA sequencing analysis using primers designed to bind to the pHEC1 plasmid upstream and downstream of the multiple cloning site. The EcProRS gene in plasmid, pHC19 and in the final expression plasmid, pHC20, is under the control of the constitutive P_L *tet* promoter. The bacterial tRNA^{Pro} was constructed by annealing oligonucleotide primers containing the entire EctRNA^{Pro} gene sequence. The primers were designed to incorporate *Eco*R I and *Pst* I sticky end, 4 nucleotide base overhangs for direct cloning of the tRNA gene into the acceptor plasmid, pSU81.

EcotRNAPro-F

5′ - AATTCCGGCACGTAGCGCAGCCTGGTAGCGCACCGTCATGGGGTG

TCGGGGGTCGGAGGTTCAAATCCTCTCGTGCCGACCACTGCA-3′

EcotRNAPro-R

5′- GTGGTCGGCACGAGAGGATTTGAACCTCCGACCCCCGACACCCCATG

ACGGTGCGCTACCAGGCTGCGCTACGTGCCGG-3′

The forward and reverse primers corresponding to the EctRNA^{Pro} sequence were annealed using a BioRad MJ Mini thermocycler (Hercules, CA). Annealing was carried out by dissolving the primers in distilled, deionized water (ddH_2O) to a final

concentration of 0.5 μg/μL. Ten μL of each of the two primers were mixed together with 4 μL 5 M NaCl, 4 μL $MgCl_2$, and 172 μL ddH_2O. By gradually decreasing the temperature of the reactions from 99 °C to 30 ºC (decreased by 5 °C every 3 min), the DNA strands were annealed together. The double-stranded DNA fragment was visualized by DNA agarose gel electrophoresis (4 % NuSieve agarose). The tRNA gene DNA was phosphorylated at the 5′ and 3′ positions using a T4 polynucleotide kinase enzyme. The pSU81 plasmid was double-digested with *Eco*R I/*Pst* I restriction endonuclease enzymes and the EctRNA gene DNA was ligated into the pSU81 plasmid to generate plasmid, pHC12. The sequence of the EctRNA sequence was confirmed by automated DNA sequencing analysis.

The pSU81 plasmid, containing the $EctRNA^{Pro}$ flanked by the *lpp* tRNA promoter and *rrnC* tRNA terminator sequences, was double-digested with *Nhe* I/*Pvu* I restriction endonucleases to isolate the tRNA expression cassette. The pHC19 plasmid, already containing the EcProRS gene, was also double-digested with *Nhe* I/*Pvu* I, and the tRNA cassette was ligated into the plasmid to generate plasmid pHC20. The sequences of the EcProRS and $EctRNA^{Pro}$ genes were confirmed by automated DNA sequencing analysis.

Other aminoacyl-tRNA synthetase genes

***Thermus thermophilus* prolyl-tRNA synthetase**

The prolyl-tRNA synthetase gene was amplified from the *T. thermophilus* genome using gene specific primers that were designed to introduce *Bam*H I and *Pst* I restriction sites at the 5′ and 3′ termini of the TtProRS gene. After amplification, the gene product was double-digested with *Bam*H I/*Pst* I and the DNA was purified by gel isolation. The TtProRS was then ligated into the *Bam*H I/*Pst* I double-digested acceptor

plasmid, pQE-80L. The sequence of the TtProRS gene was confirmed by automated DNA sequencing analysis using primers designed to bind to the pQE-80L plasmid upstream and downstream of the multiple cloning site.

***Saccharomyces cerevisiae* prolyl-tRNA synthetase**

The prolyl-tRNA synthetase gene was amplified from the *S. cerevisiae* genome using gene specific primers that were designed to introduce *Kpn* I and *Xba* I restriction sites at the 5′ and 3′ termini of the ScProRS gene. After amplification, the gene product was double-digested with *Kpn* I/*Xba* I and the DNA was purified by gel isolation. The ScProRS was then ligated into the *Kpn* I/*Xba* I double-digested acceptor plasmid, pHEC1. The sequence of the ScProRS gene was confirmed by automated DNA sequencing analysis.

Confirming orthogonality

Construction of "mix and match" plasmids

The MjProRS/MjtRNA^{Pro} pair was cloned into the pHEC1 plasmid to create plasmid pHC15 as described previously. The EcProRS/EctRNA^{Pro} pair was cloned into the pHEC1 plasmid to create plasmid pHC20 as described previously. The EctRNA^{Pro} gene was excised from the pHC12 plasmid by double-digestion with *Nhe* I/*Pvu* I. The tRNA gene DNA and the *Nhe* I/*Pvu* I double-digested acceptor plasmid, pHC10 already containing a copy of the MjProRS gene, were purified by gel isolation. The tRNA gene cassette was ligated into the pHC10 plasmid to create the MjProRS/EctRNA^{Pro} pair construct in plasmid, pHC17. The MjtRNA^{Pro} gene was excised from the pHC13 plasmid by double-digestion with *Nhe* I/*Pvu* I. The tRNA gene DNA and the *Nhe* I/*Pvu* I double-digested acceptor plasmid, pHC19 already containing a copy of the EcProRS gene, were

purified by gel isolation. The tRNA gene cassette was ligated into the pHC19 plasmid to create the MjProRS/EctRNA^{Pro} pair construct in plasmid, pHC21.

Functional Assay

UQ27 ProRS $^{ts, ProA}$ cells were transformed with appropriate (see results) plasmids containing prolyl-tRNA synthetase genes of interest. A single colony, for each cell type, was picked from the transformation plate with a sterile inoculation loop. The colony was spread onto a new plate using the spiral method or by streaking. UQ27 ProRS $^{ts, ProA}$ cells were plated on solid LB media supplemented with kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL). Expression from the synthetase containing plasmids was under the control of the constitutive P_L *tet* promoter so the plates were not supplemented with IPTG. Cells were grown in triplicate, overnight on the plates at 30 °C, 37 °C, and 42 ºC, and the results were recorded the following day.

Protein Expression and Purification

Small-scale expression in the UQ27$^{ts, ProA}$ strain*

pHC8 plasmid DNA (pQE-80L containing the **elastin-CCC** gene) was co-transformed with pWK1 plasmid DNA (containing a recombinant, wild-type *E. coli* prolyl-tRNA synthetase gene) or the empty plasmid, pHEC1 (lacking the synthetase/tRNA pair), into the UQ27$^{ts, ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). pHC8 plasmid DNA also was also co-transformed with pHC15 plasmid DNA (containing the recombinant, wild-type *M. jannaschii* prolyl-tRNA synthetase/tRNA^{Pro} gene), or other appropriate plasmids, into the UQ27$^{ts, ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol

(34 μg/mL), and ampicillin (100 μg/mL). Five mL LB media, supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL), were inoculated with single colonies of $UQ27^{ts,\ ProA}$ cells, harboring the appropriate plasmids, and grown overnight at 30 °C on a rotator. Two mL of the overnight cultures were transferred to 48 mL LB (supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL) that had been added to 250 mL Erlenmeyer flasks for a total volume of 50 mL culture.

The 50 mL cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), protein expression was induced with IPTG. For induction, 50 uL, 1M IPTG were added to the flasks for a final concentration of 1 mM IPTG. The cultures were incubated at 42 °C for 3 h. OD_{600} readings were monitored for each of the cultures. 1 mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 μL ddH_2O and stored at –20 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 μL sample culture, 12.5 μL 2 x sample buffer, 1 μL DTT (dithiothreitol), and 6.5 μL ddH_2O for a total of 25 μL. Samples were placed at 95 ºC for 5 min.

*For expression in cell lines other than the UQ27 strain, the only modification from the outlined procedure is that cells were grown and incubated before and after induction at 37 ºC rather than at 30 ºC and then 42 °C.

Small-scale expression in minimal media for proline analogue incorporation

pHC8 plasmid DNA (pQE-80L containing the **elastin-CCC** gene) was co-transformed with pWK1 plasmid DNA (containing a recombinant, wild-type *E. coli* prolyl-tRNA synthetase gene) or the empty plasmid, pHEC1 (lacking the synthetase/tRNA pair), into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). pHC8 plasmid DNA also was also co-transformed with pHC15 plasmid DNA (containing the recombinant, wild-type *M. jannaschii* prolyl-tRNA synthetase/$tRNA^{Pro}$ gene), or with other appropriate plasmids, into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). Five mL LB media, supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL), were inoculated with single colonies of $UQ27^{ts,\ ProA}$ cells, harboring the appropriate plasmids, and grown overnight at 30 °C on a rotator. Two mL of the overnight cultures were transferred to 48 mL NMM + proline or SMM + proline liquid media (supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL) that had been added to 250 mL Erlenmeyer flasks for a total volume of 50 mL growth culture.

The 50 mL cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), the 50 mL cultures were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles to spin down the cells. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge

bottles. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were finally resuspended in 9 mL of NMM - proline or SMM - proline minimal media for each of 2 small duplicate flasks (18mL NMM or SMM - proline total) for each cell type. The cell cultures were incubated at 42 °C for 30 min with shaking at 225 rpm to deplete the remaining cellular levels of proline.

For expressions involving proline analogues, the osmolarity of the culture was adjusted prior to induction via addition of the appropriate osmolytes to a final concentration of either 600 mM for NaCl or 800 mM for sucrose. For example, 1 mL of a 5 M NaCl stock solution (29.22 g NaCl in 100 mL H_2O, sterilized) was added to each expression flask. The osmolarity of the culture was adjusted for incorporation of all analogues with the exceptions of (2*S*,4*R*)-4-fluoroproline and (2*S*,4*S*)-4-fluoroproline which did not require osmotic shock for successful incorporation. In addition, proline analogues, such as 2*S*,4*R*)-4-hydroxyproline or (2*S*,4*R*)-4-fluoroproline, were prepared as 1 mL 200 mM stock solutions in PBS at pH 7.4 and stored at 4 °C. L-proline was prepared as a 1 mL 200 mM stock solution in PBS at pH 7.4 and stored at 4 °C. A 100 µL aliquot of the L-proline or the proline analogue stock solutions was added to the 10 mL cell cultures for a final concentration of 2 mM proline or proline analogue. To induce protein expression, 10 µL 1M IPTG were added to each flask for a final

concentration of 1 mM IPTG. The IPTG 1000x (1M) stock was prepared by dissolving 596 mg IPTG in 2.5 mL ddH_2O followed by filter sterilization.

The cultures were incubated at 42 °C with shaking at 225 rpm for 3 h after induction. OD_{600} readings were monitored for each of the cultures. One mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 µL ddH_2O and stored at –20 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 µL sample culture, 12.5 µL 2 x sample buffer, 1 µL DTT (dithiothreitol), and 6.5 µL ddH_2O for a total of 25 µL. Samples were placed at 95 ºC for 5 min.

Large-scale expression

pHC8 plasmid DNA (pQE-80L containing the **elastin-CCC** gene) was co-transformed with pWK1 plasmid DNA (containing a recombinant, wild-type *E. coli* prolyl-tRNA synthetase gene) or the empty plasmid, pHEC1 (lacking the synthetase/tRNA pair), into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL). pHC8 plasmid DNA also was also co-transformed with pHC15 plasmid DNA (containing the recombinant, wild-type *M. jannaschii* prolyl-tRNA synthetase/$tRNA^{Pro}$ gene) into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL). Five mL LB media, supplemented with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL) were inoculated with single colonies of $UQ27^{ts,\ ProA}$ cells, harboring the appropriate plasmids, and grown overnight at 30 °C on a rotator. Five mL of the overnight cultures were transferred to 495 mL LB

(supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL) that had been added to 2,800 mL Erlenmeyer flasks for a total volume of 500 mL culture. Two flasks were used to express 1000 mL (1L) total volume of culture.

The large cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), protein expression was induced with IPTG. For induction, 500 uL, 1M IPTG were added to the flasks for a final concentration of 1 mM IPTG. The cultures were incubated at 42 °C for 4 h. OD_{600} readings were monitored for each of the cultures. 1 mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 uL ddH_2O and stored at –20 °C. After 3 h of induction, the two large cultures (1 L total volume) were spun down at 4,000 x g for 20 min. The cell pellets were resuspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and stored at –80 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 μL sample culture, 12.5 μL 2 x sample buffer, 1 μL DTT (dithiothreitol), and 6.5 μL ddH_2O for a total of 25 μL. Samples were placed at 95 ºC for 5 min.

Purification of elastin-CCC protein

Purification of the **elastin-CCC** protein was facilitated by a C-terminal decahistidine tag. A large-scale expression was carried out to obtain a large amount of the **elastin-CCC** protein in the expression strain, $UQ27^{ts,\ ProA}$. The frozen cells were lysed by three freeze/thaw cycles. Lysozyme (1 mg/mL), EDTA-free protease inhibitor cocktail, benzonase (25 units/mL), and $MgCl_2$ (1 mM) were added to the lysate and the mixture was incubated with shaking (225 rpm) at 4 °C, overnight. The cell lysate was

centrifuged at 10,000 x g for 30 min at 4 °C. The supernatant and pellet were separated and analyzed by SDS-PAGE to determine the location of the target protein. SDS-PAGE analysis indicated that the majority of the target protein was present in the soluble fraction with some residual protein in the pellet. For the soluble elastin-mimetic protein, the supernatant was loaded onto TALON® Co^{2+} metal affinity resin (5 mL) and washed with lysis buffer (50 mL) containing 20 mM imidazole. The target protein was eluted with elution buffer (20 mL, 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.0) and dialyzed (MWCO = 10 kDa) against distilled deionized water (5 × 4 L). The dialysate was lyophilized to produce a white spongy solid.

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS)

The molar mass of **elastin-CCC** was determined by MALDI-TOF MS on an Applied Biosystems Voyager System 428 mass spectrometer in the positive linear mode. The matrix, 2-(4-hydroxyphenylazo)benzoic acid (HABA), was used at a concentration of 10 mg/ml in a 50:50 mixture of water and 2-propanol. The protein solution (1 mg/ml in ddH_2O) was mixed with the matrix solution in a ratio of 1:10 and dried under vacuum or air. Bovine serum albumin was used as a standard for external calibration. Sample preparation was repeated in triplicate and the average mass was calculated for each of 3 trials of MALDI-TOF-MS mass analysis.

Primers used in this study

Lar-F: 5′-CGAATCCCGGGCGATCGCGGAGATTTCCTGGAAGATGCCAGG-3′

Lar-R: 5′-AGCTACCCGGGCTAGCCCGTTCGTAAGCCATTTCCGCTCGC-3′

tRNAPro-F

5′-AATTCGGCTCGTTGGTCTAGGGGTATGATTCTCGCTTGGGGTGCGAG

AGGTCCCGGGTTCAAATCCCGGACGAGCCCCCACTG-3′

tRNAPro-R

5′-GTGGGGGCTCGTCCGGGATTTGAACCCGGGACCTCTCGCACCCCAA

GCGAGAATCATACCCCTAGACCAACGAGC-3′

MjtRNAPro-F

5′ - AATTCGGGGCCGTGGGGTAGCCTGGATATCCYGTGCGCTTGGGGGGC

GTGCGACCCGGGTTCAAGTCCCGGCGGCCCCACCACTGCA-3′

MjtRNAPro-R

5′- GTGGTGGGGCCGCCGGGACTTGAACCCGGGTCGCACGCCCCCCAAG

CGCACRGGATATCCAGGCTACCCCACGGCCCCG-3′

EcotRNAPro-F

5′ - AATTCCGGCACGTAGCGCAGCCTGGTAGCGCACCGTCATGGGGTG

TCGGGGGTCGGAGGTTCAAATCCTCTCGTGCCGACCACTGCA-3′

EcotRNAPro-R

5′- GTGGTCGGCACGAGAGGATTTGAACCTCCGACCCCCGACACCCCATG

ACGGTGCGCTACCAGGCTGCGCTACGTGCCGG-3′

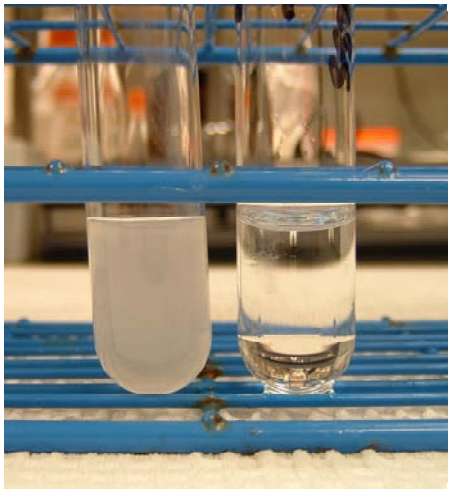
Strain	Genotype	Comments
DG99	F^{-}, *thi*-1, *endA*1, *supE44*, *hsdR*17, (*proC*::Tn10, *lacIq*, *_lacZ) M15*	Proline auxotroph
CAG18515	F^{-}, *proA*3096::Tn10Kan, *rph*-1	Proline auxotroph
UQ27	*proS*127(ts), *lacZ*4,lam^{-}, *argG*75	Proline auxotroph with a temperature sensitive ProRS
JM108	*end*A1, *gyr*A96, *hsd*R17(r_K^- m_K^+), *rec*A1, *rel*A1, *sup*E44, *thi*-1, (lac-proAB)	Proline auxotroph
Rosetta™ (Novagen, Madison, WI)	F$^-$ *ompT hsdS*B(r_B^- m_B^-) *gal dcm* pRARE (argU, argW, ileX, glyT, leuW, proL) (Cam^R)	The strain contains a plasmid that harbors rarely used *E. coli* tRNA genes for expression of eukaryotic proteins that contain rarely used codons
Rosetta Blue™ (Novagen, Madison, WI)	*endA1 hsdR17* (r_{K12}^- m_{K12}^+) *supE44 thi-1 recA1 gyrA96 relA1 lac* F′ [*proA*$^+$*B*$^+$ *lacIqZDM15*::Tn*10* (tet^R)] pRARE (argU, argW, ileX, glyT, leuW, proL) (Cam^R)	The strain contains a plasmid that harbors rarely used *E. coli* tRNA genes as well as *endA, recA,* and *lacIq mutations*

Table 2. *E. coli* cell strains used for protein expression

Results and Discussion

Preparation of the Bacterial Host Strain

Expression of the **elastin-CCC** protein was carried out in an *E. coli* bacterial host. The UQ27 strain (Genotype: *proS*127(ts), *lacZ*4,lam–, *argG*75)[32] was a target for use as a host for the expression of the protein in the presence of the orthogonal translational system. The bacteria express a temperature sensitive mutant form of the endogenous prolyl-tRNA synthetase (ProRS) enzyme (W375R mutation in the endogenous EcProRS). At 30 ºC, the bacteria express an active ProRS enabling incorporation of proline or proline analogues in host cell proteins. However, at 42 °C, the mutant ProRS is inactive and does not allow for imino acid incorporation. Transduction, using P1 donor phage, was used to confer a proline auxotrophic phenotype by homologous replacement of the *pro*A gene with an antibiotic resistance gene, *pro*A3096::Tn10 Kan, cassette. After P1 transduction, cells did not grow in media lacking proline. The UQ27^{ProA} strain, a proline auxotroph, was used for subsequent expression experiments.



Proline + Proline -

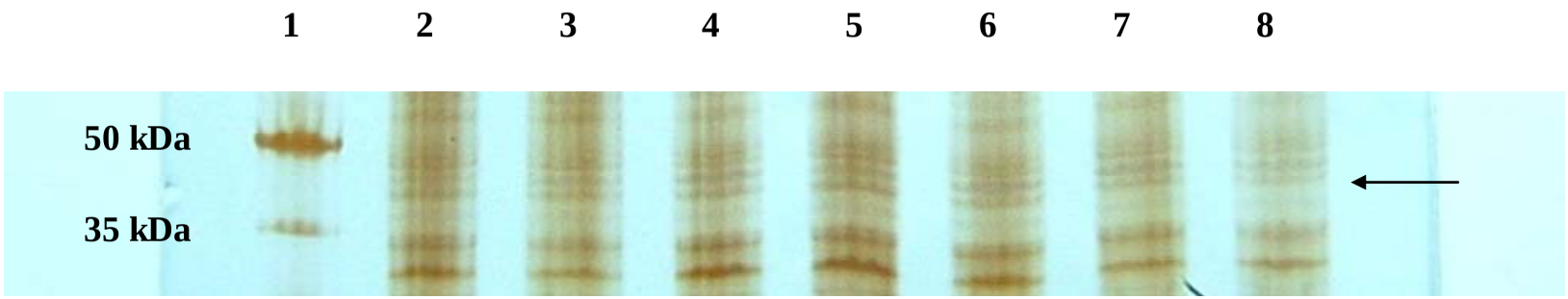
Figure 2. Transduction using P1 donor phage was used to confer a proline auxotrophic phenotype in the *E. coli* UQ27 ProRS^{ts} strain by homologous replacement of the *proA* gene with an antibiotic resistance gene *proA*3096::Tn10 Kan cassette. Cells were not able to grow without addition of proline to the growth media.

Protein Expression

Homo sapiens prolyl-tRNA synthetase/tRNA (ProRS/tRNA^{Pro}) pair

Initial efforts were focused on using a *H. sapiens* orthogonal translational system for analogue incorporation and **elastin-CCC** expression in *E. coli*. The human pair was not functional in the *E. coli* host organism (Fig. 3). Small-scale expressions of the human pair with co-expression of the **elastin-CCC** gene (pHC8 and pHC9, see Chapter 2) in the UQ27^{ts,ProA} in LB media, did not give rise to detectable protein production in SDS-PAGE analysis. Expression of a human orthogonal pair in *E. coli* would be ideal for production of human genes such as elastin and collagen, but a human synthetase and tRNA may need folding elements not present in a bacterial expression host. Although the pair would likely be truly orthogonal and no cross-reaction with *E. coli* genetic elements would be expected, the synthetase enzyme did not appear to be functional in the host possibly due to a folding problem in the prokaryotic *E. coli* host organism.

Expression of **elastin-CCC** was confirmed in the Rosetta™ and Rosetta Blue™ strains to rule out the possibility that the gene was defective. The Rosetta™ strains harbor a plasmid for expression of the *pro*L tRNAPro gene encoding the GGG anticodon in the host. Expression of the **elastin-CCC** gene was observed due to the higher levels of the usually rare *pro*L tRNA. Expression of **elastin-CCC** was visible after 3-4 h induction in the Rosetta™ and Rosetta Blue™ strains (Fig. 4).



1. Perfect Protein Marker

2. pHC8 (**elastin-CCC**), before induction

3. pHC8, uninduced

4. pHC8, induced

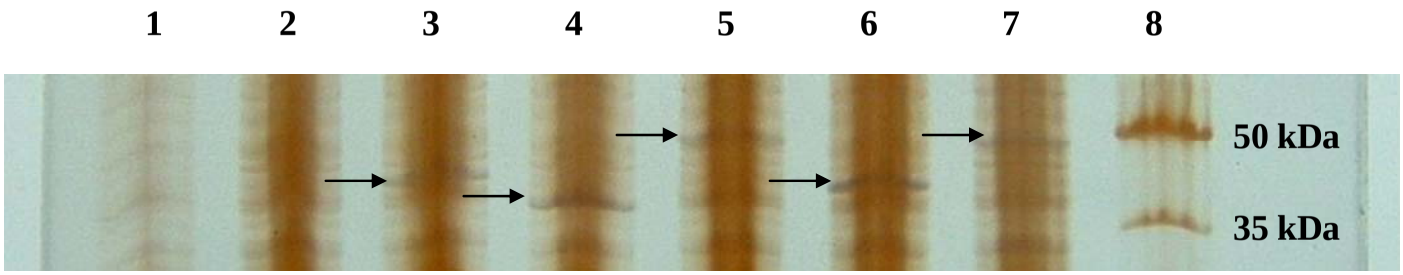
5. pHC8 + pHC2 (human pair), uninduced

6. pHC8 + pHC2 (human pair), induced

7. pHC8 + pHC1 (Human ProRS), uninduced

8. pHC8 + pHC1 (Human ProRS), induced

Figure 3. SDS-PAGE analysis with silver stain indicating expression of **elastin-CCC** in the presence of the human prolyl-tRNA synthetase and the human ProRS/tRNAPro pair. No distinctive protein expression band was visible after 3-4 h induction in LB or minimal media. The **elastin-CCC** band should appear around 45 KDa.



1. pHC8 (**elastin-CCC**), before induction, Rosetta™
2. pHC8, uninduced, Rosetta™ strain
3. pHC8 (clone 1), induced, Rosetta™ strain
4. pHC8 (clone 2), induced, Rosetta™ strain
5. pHC9, induced, Rosetta™ strain
6. pHC8 (clone 1), induced, Rosetta Blue™
7. pHC9, induced, Rosetta Blue™
8. Perfect protein marker

Figure 4. SDS-PAGE analysis with silver stain indicating expression of **elastin-CCC** in the Rosetta™ and Rosetta Blue™ *E. coli* strains. Distinctive protein expression bands were visible after 3-4 h induction in LB. The **elastin-CCC** bands appear around 45 kDa. The pHC8 plasmid harbors an **elastin-CCC** gene that is ~1500 bp in length while the pHC9 plasmid harbors an **elastin-CCC** gene that is ~1800 bp in length.

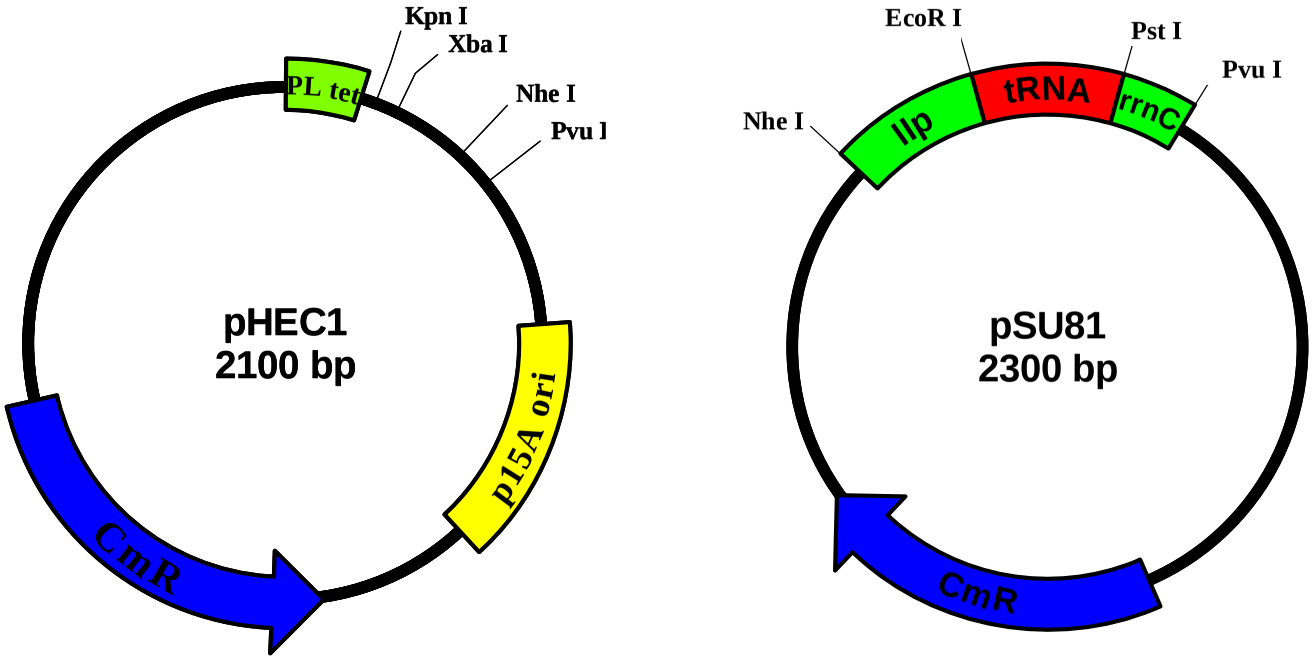


Figure 5. The pHEC1 plasmid is derived from the pPROTet/Lar plasmid. The pHEC1 plasmid was generated by inverse PCR amplification with primers designed to incorporate the *Nhe* I and *Pvu* I restriction endonuclease sites into the plasmid to enable cloning of tRNA cassettes in the plasmid. The multiple cloning site in the plasmid also contains the *Kpn* I and *Xba* I restriction endonuclease sites for cloning ProRS genes into the plasmid. The plasmid is a highly useful vector for cloning and expression of orthogonal aminoacyl-tRNA synthetase/tRNA pairs in *E. coli*. Plasmid pSU81 is also a useful plasmid for cloning orthogonal pairs as this plasmid enables cloning of tRNA genes (hypothetical tRNA gene shown here) as a cassette with flanking tRNA promoter and terminator sequences for optimal and constitutive expression of tRNA genes in an *E. coli* host.

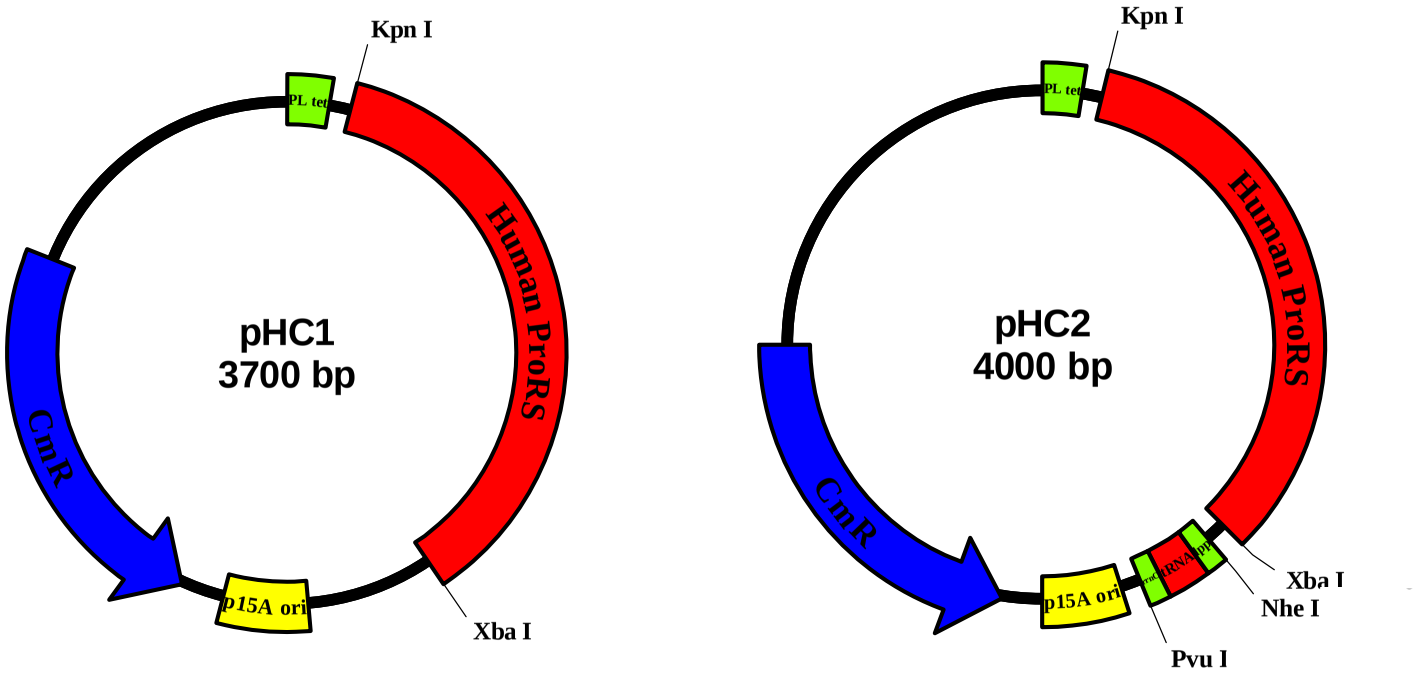
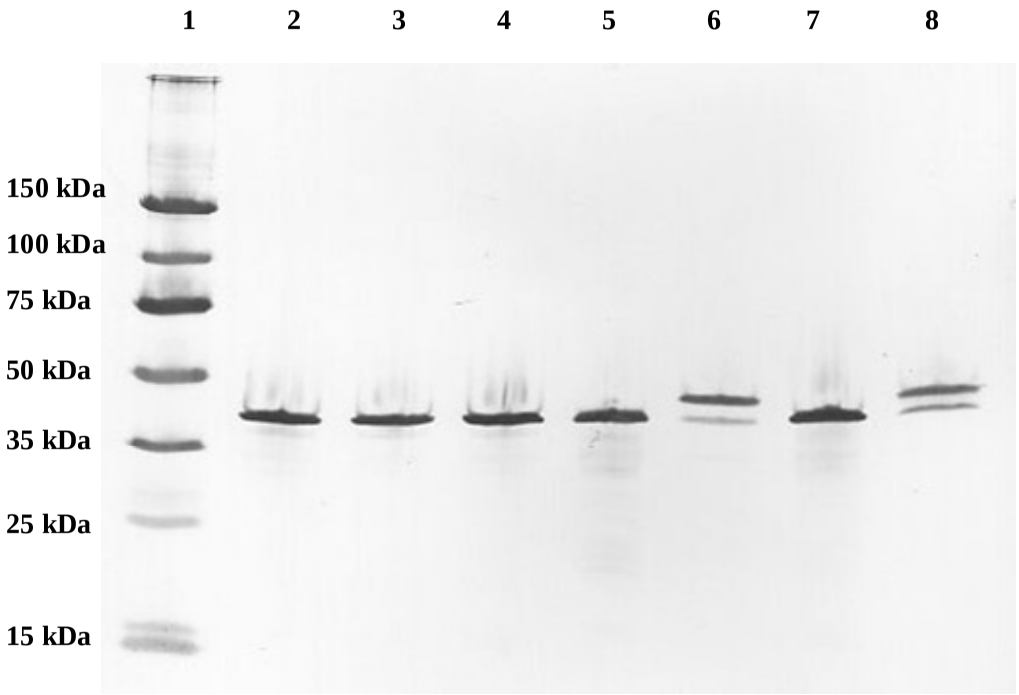


Figure 6. The pHC1 plasmid is derived from the pHEC1 plasmid and harbors the gene for HsProRS. The pHC2 plasmid is derived from the pHC1 plasmid and has the HsProRS gene with the $HstRNA^{Pro}$ gene cassette.

***Methanococcus jannaschii* prolyl-tRNA synthetase/tRNA (ProRS/tRNA^{Pro}) pair**

The **elastin-CCC** protein was expressed in *E. coli* strains CAG18515 and UQ27$^{ts, ProA}$. The UQ27$^{ts, ProA}$ was found to be optimal for the expression of the elastin-mimetic protein in the presence of the orthogonal pair as some background level expression was seen in the CAG18515 strain (compare Fig. 7 and Fig. 8). Although both strains are proline auxotrophs and could be useful for analogue incorporation, the CAG18515 strain has a functional copy of the EcProRS encoded by the genomic DNA. It is suspected that this endogenous synthetase enables expression of **elastin-CCC** even without the MjProRS/MjtRNA^{Pro} pair as the endogenous synthetase is able to charge the rare, yet endogenous, EctRNA^{Pro} with proline and enable translation through the CCC codons of the **elastin-CCC** mRNA. The UQ27$^{ts, ProA}$ strain, however, has a temperature sensitive copy of the endogenous synthetase and without the function of the endogenous synthetase at 42 °C, protein expression is carried out by the translational machinery in the host as well as the orthogonal function of the MjProRS/MjtRNA^{Pro} pair. A recombinant copy of the EcProRS (no tRNA gene added) enables **elastin-CCC** expression giving evidence for the need for using the temperature sensitive, UQ27$^{ts, ProA}$ strain (Fig. 8).



1. Perfect protein marker

2. pHC8 + pHEC1 (no pair), 3 h after induction

3. pHC8 + pWK1 (EcProRS), 3 h after induction

4. pHC8 + pHC15 (Mj Pair), 3 h after induction

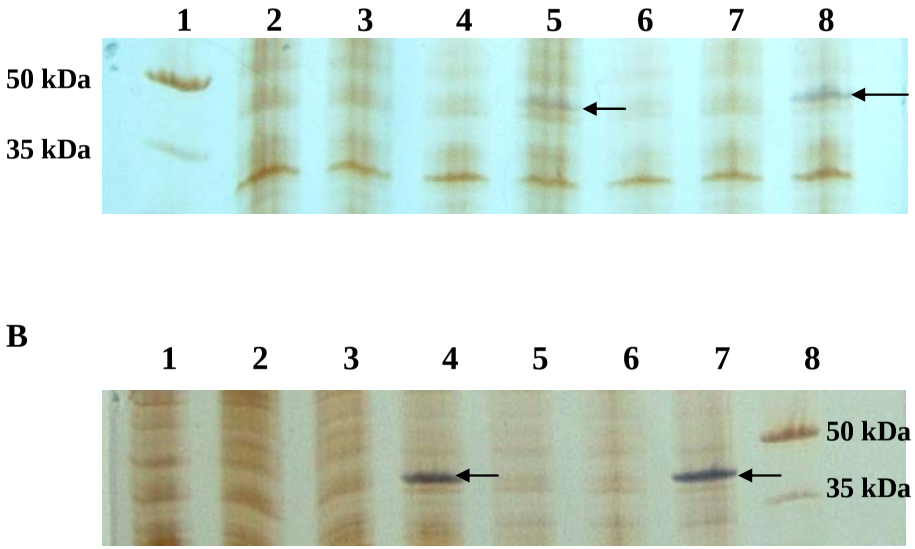
5. pHC8 + pHEC1 (no pair), + proline

6. pHC8 + pHEC1 (no pair), + hydroxyproline

7. pHC8 + pHC15 (Mj pair), + proline

8. pHC8 + pHC15 (Mj pair), + hydroxyproline

Figure 7. Western blot analysis indicating expression of **elastin-CCC** in the presence of the MjProRS/MjtRNA^{Pro} pair in the CAG18515 *E. coli* strain (Lanes 4, 7, 8). Background-level expression of the protein is seen in the absence of the pair (Lanes 2, 5, 6) due to the endogenous expression of a functional EcProRS enzyme. Expression of the EcProRS alone also enables expression of **elastin-CCC** (Lane 3). With expression in minimal media, incorporation of (2S, 4R)-4-hydroxyproline is evident (Lanes 7 and 8) with the Mj pair present. However, the CAG18515 strain allows for expression without the pair present as well (Lanes 5 and 6). Also, there are slight bands visible for mixed proline/hydroxyproline incorporation in addition to the correct bands for hydroxyproline incorporation possibly due to the activity of the endogenous EcProRS in this strain scavenging proline (Lanes 6 and 8).



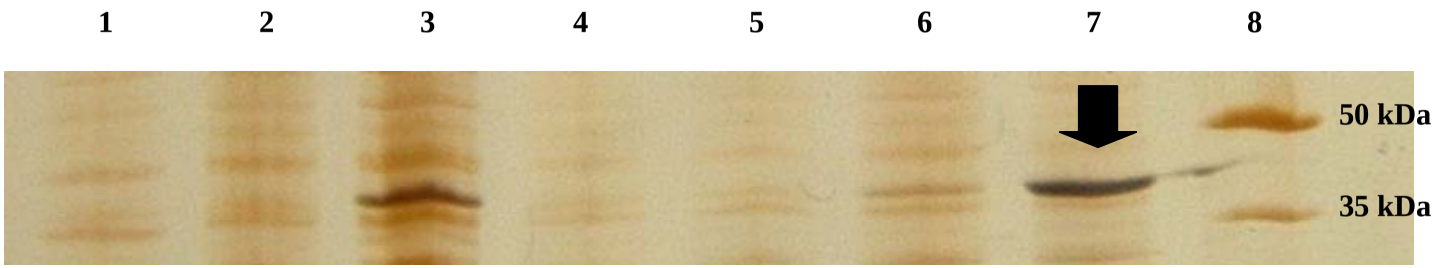
A

1. Perfect protein marker
2. pHC8 (elastin-CCC), after 3 h, uninduced
3. pHC8 (elastin-CCC), 3 h after induction
4. pHC8 + pHEC1 (no pair), 3 h after induction
5. pHC8 + pWK1 (EcProRS), 3 h after induction
6. pHC8 + pHC2 (Human pair), 3 h after induction
7. pHC8 + pHC15 (Mj Pair), after 3 h, uninduced
8. pHC8 + pHC15 (Mj Pair), 3 h after induction

B

1. pHC8 + pHC15, before induction
2. pHC8 (elastin-CCC), 5 h after induction
3. pHC8 + pHC2 (Human pair), 5 h after induction
4. pHC8 + pHC15 (Mj Pair), 5 h after induction
5. pHC8 (elastin-CCC), 24 h after induction
6. pHC8 + pHC2 (Human pair), 24 h after induction
7. pHC8 + pHC15 (Mj Pair), 24 h after induction
8. Perfect protein marker

Figure 8. SDS-PAGE analysis with silver stain indicating expression of **elastin-CCC** in the presence of the MjProRS/MjtRNA^{Pro} pair in LB in the UQ27$^{ts, ProA}$ strain at 42 ºC. The Mj pair enabled protein expression while the human pair did not (**A**. Compare lanes 6 and 8). The EcProRS alone enables some protein expression (**A**. Lane 5) which gives evidence for the need for using the UQ27$^{ts, ProA}$ strain to eliminate detectable background level expression coming from the endogenous EcProRS. **Elastin-CCC** expression appears to be stable and consistent even 5 and 24 h after induction (**B**).



1. pHC8 (**elastin-CCC**) + pHC15 (Mj pair), before induction

2. Negative control, pHC8 + pHEC (no pair)

3. pHC8 + pHC19 (EcProRS only)

4. pHC8 + pHC10 (MjProRS only, no MjtRNA)

5. pHC8 + pHC13 (MjtRNA only, no MjProRS)

6. pHC8 + pHC15 (Mj Pair), uninduced

7. pHC8 + pHC15, induced

8. Perfect protein marker

Figure 9. SDS-PAGE analysis with silver stain indicating expression of **elastin-CCC** in the presence of the MjProRS/MjtRNA pair. Co-expression of the MjProRS or MjtRNA alone does not result in protein expression. The need for expressing both members of the pair simultaneously, suggests that the pair functions independently of the endogenous translational machinery and there is no cross-charging between the orthogonal pair and the *E. coli* machinery.

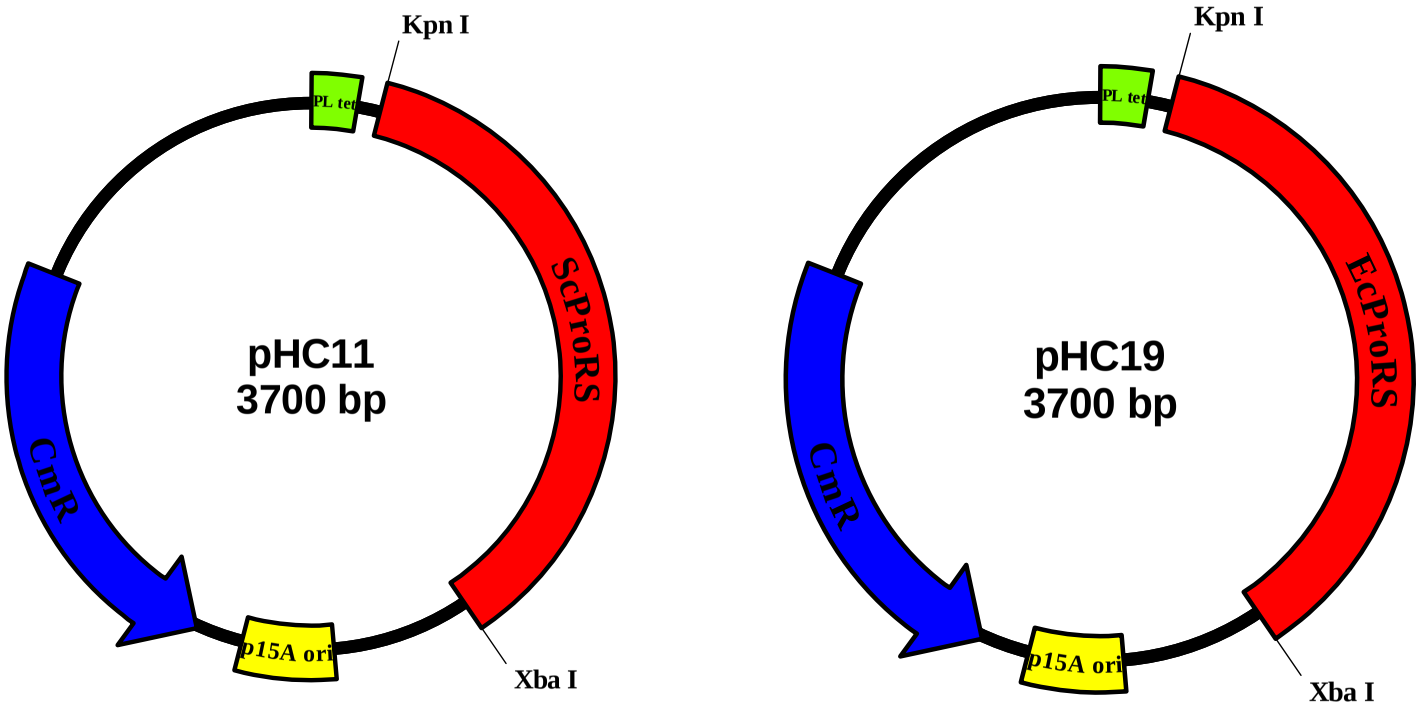


Figure 10. The pHC11 plasmid is derived from the pHEC1 plasmid. A recombinant copy of the ScProRS gene is encoded by the plasmid for constitutive protein expression in a bacterial host. The pHC19 plasmid (same as pWK1) is derived from the pHEC1 plasmid. A recombinant copy of the EcProRS gene is encoded by the plasmid for protein expression or for cloning purposes.

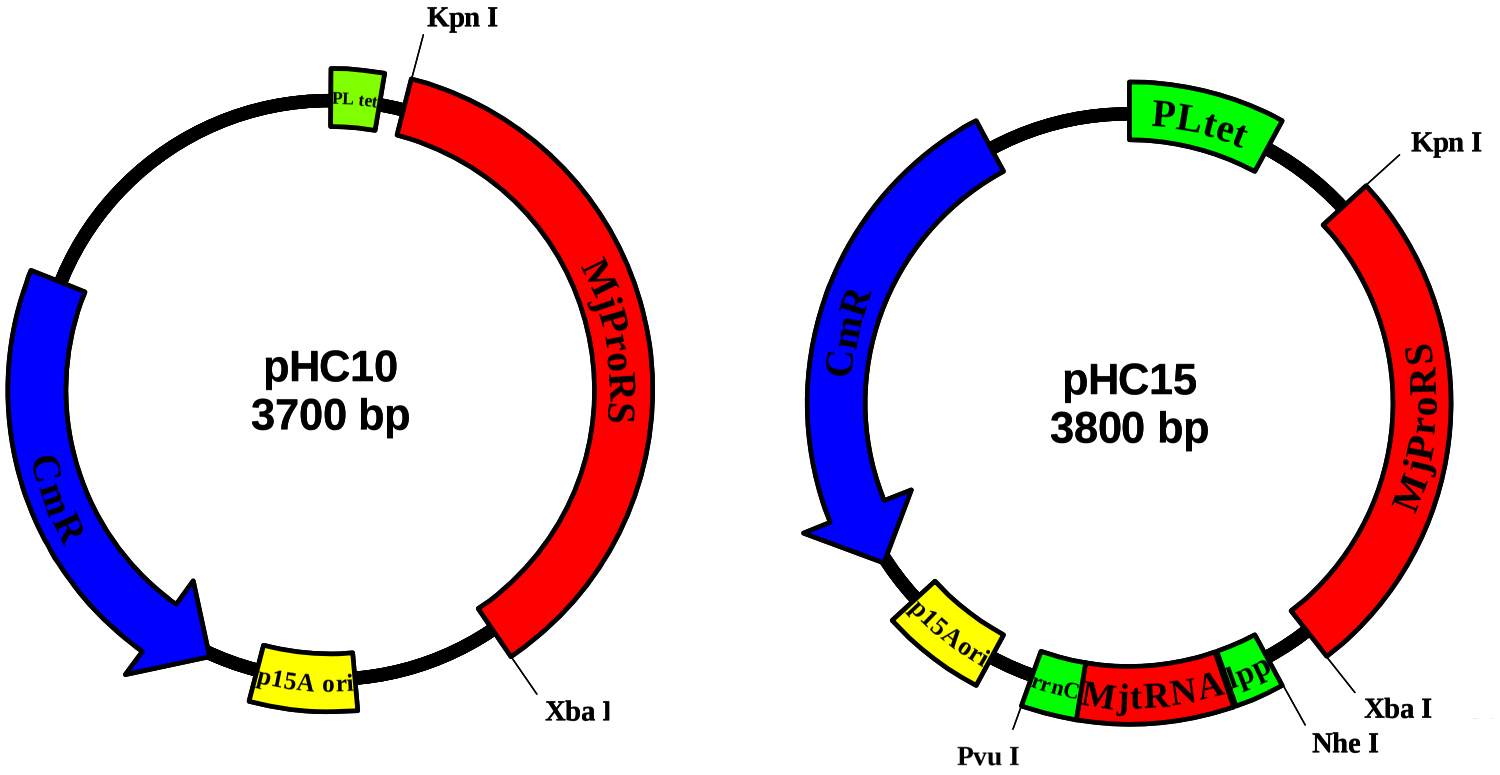


Figure 11. Plasmid pHC10 is derived from the pHEC1 plasmid. A recombinant copy of the MjProRS gene is encoded by the plasmid for constitutive protein expression in a bacterial host. The pHC15 plasmid, derived from pHC10, harbors the MjProRS/MjtRNA^{Pro} pair for expression of the orthogonal pair in *E. coli*.

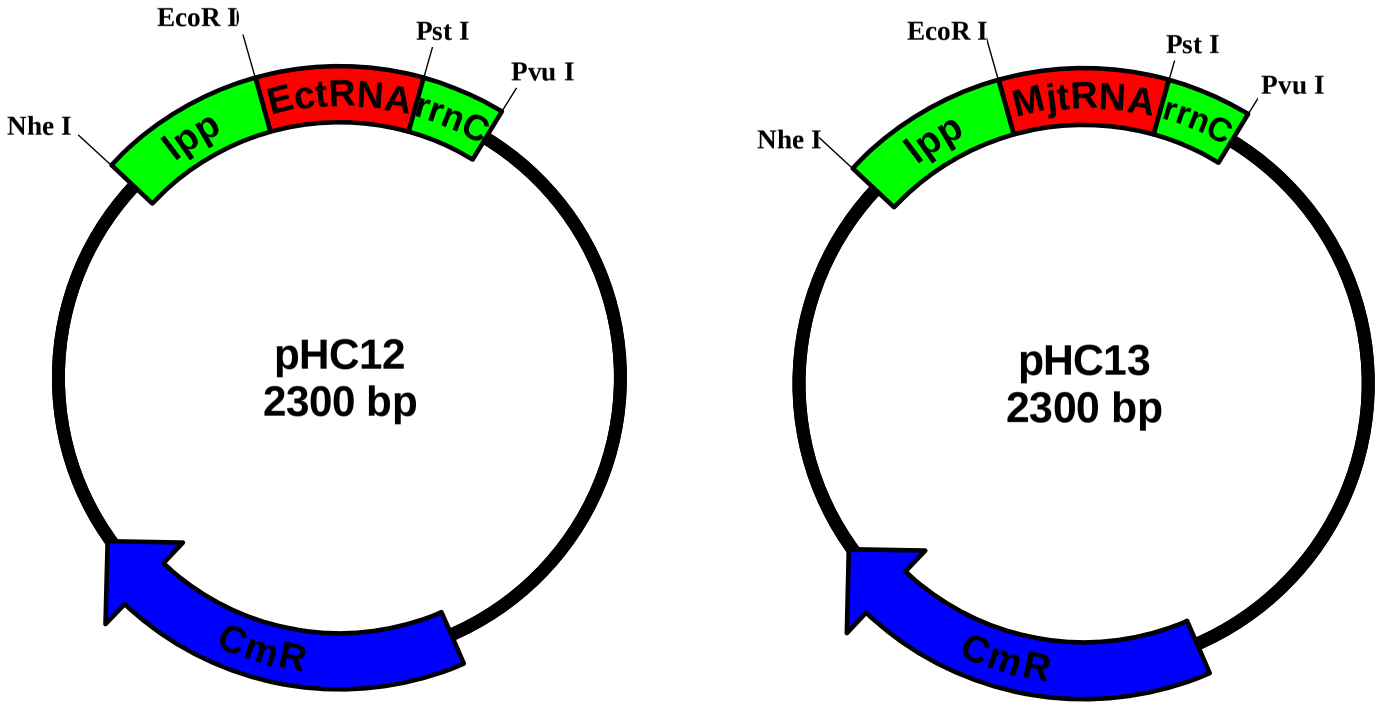


Figure 12. The pHC12 plasmid is derived from the pSU81 plasmid. The EctRNA^{Pro} gene cassette in the plasmid can be removed by double-digestion with *Nhe* I/*Pvu* I restriction endonucleases for cloning of the gene with flanking tRNA promoter and terminator sequences. The pHC13 plasmid is derived from the pSU81 plasmid. The MjtRNA^{Pro} gene cassette in the plasmid can be removed by double-digestion with *Nhe* I/*Pvu* I restriction endonucleases for cloning of the gene with flanking tRNA promoter and terminator sequences.

Expression of the MjProRS or of the MjtRNA^{Pro} alone were not sufficient for expression of **elastin-CCC** in the UQ27$^{ts,\ ProA}$ strain (Fig. 9). Without both members of the pair present, together, high-level protein expression was not possible due to the rarity of the EctRNA *pro*L gene and the inactive endogenous ProRS at 42 ºC. The pair appears to be truly orthogonal and exhibited no detectable cross-reaction with the endogenous *E. coli* translational machinery (Fig. 9). The **elastin-CCC** gene has 80 sites for incorporation of proline or proline analogues in the primary sequence, all encoded by the CCC codon. Proline comprises approximately 20 % of the overall **elastin-CCC** sequence. Protein expression, in the presence of the MjProRS/tRNA^{Pro} pair, demonstrated high-level read through and function of the MjProRS/tRNA^{Pro} pair.

"Mix and match" experiment

In order to test for orthogonality and to rule out cross-reaction between the MjProRS/MjtRNAPro pair and the endogenous *E. coli* translational machinery, plasmids were constructed for testing orthogonality in a "mix-and-match" experiment. A series of plasmids were constructed that harbored each of the components of the orthogonal pair and of the *E. coli* pair (Fig. 13). The MjProRS/MjtRNA^{Pro} was encoded by the pHC15 plasmid, while the EcProRS/EctRNA^{Pro} pair was encoded by the pHC20 plasmid. Combinations of the ProRS and tRNA genes were encoded by plasmids, pHC17 (MjProRS/EctRNA^{Pro}) and pHC21 (EcProRS/MjtRNA^{Pro}). Protein expression was evident in the presence of the Mj pair and in all cases where the EcProRS was expressed (Fig. 14). Neither the MjProRS nor the MjtRNA^{Pro} alone enabled protein expression. The MjProRS was not able to charge the EctRNA^{Pro} with proline suggesting no cross-reaction (Fig. 14). The EctRNA^{Pro} alone, without the EcProRS, does not enable protein

expression in the UQ27$^{ts,\ ProA}$ strain while the presence of the EctRNA^{Pro} in the Rosetta™ and Rosetta Blue™ strains enabled expression. It is unknown whether the EcProRS is able to charge the MjtRNA^{Pro}, but it is unlikely based on previous observations and the differences between recognition regions of the tRNA molecules in the two organisms[14]. One way to test for orthogonality would be to carry out *in vitro* aminoacylation experiments. In our system, however, it is not critical for our system even if the EcProRS is able to charge the MjtRNA, as we use inactivation of the EcProRS with expressions in the UQ27$^{ts,\ ProA}$ strain.

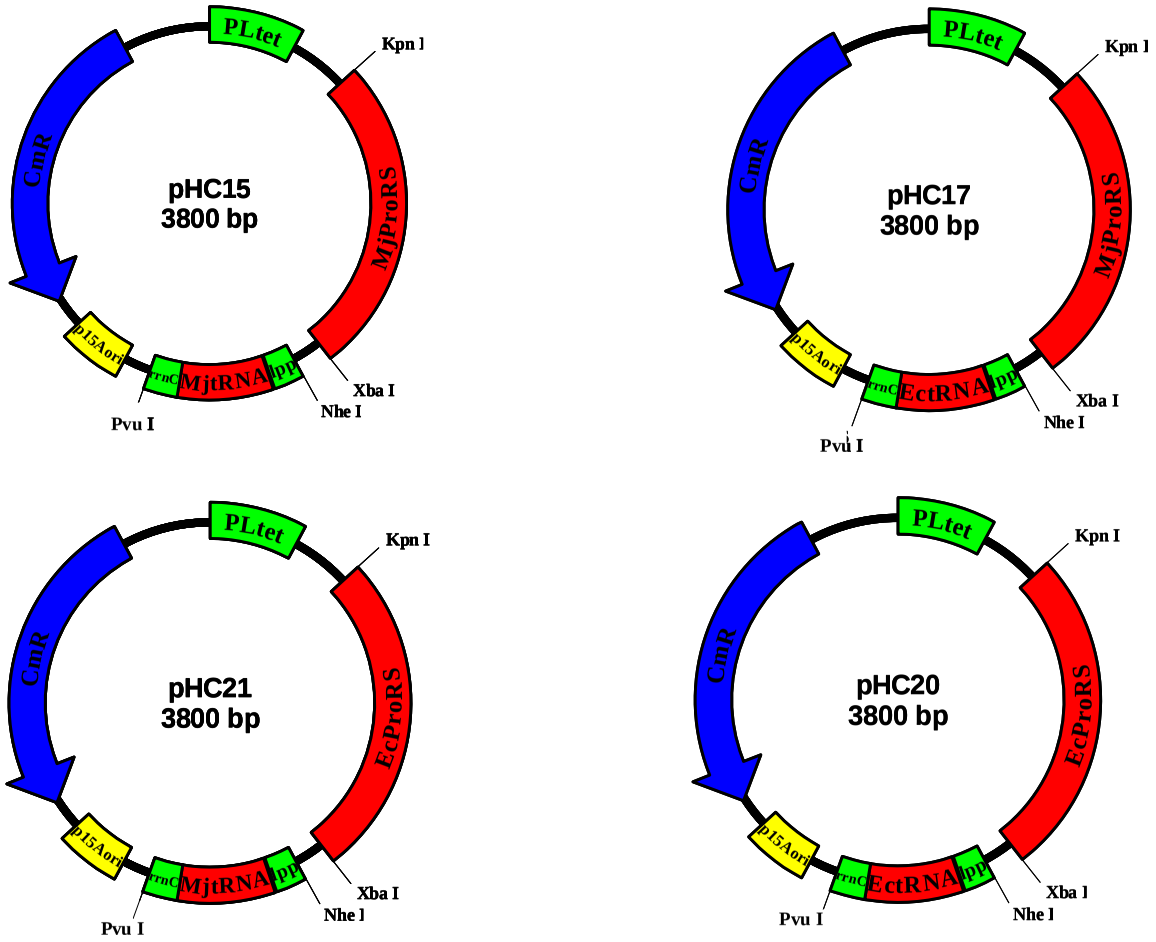
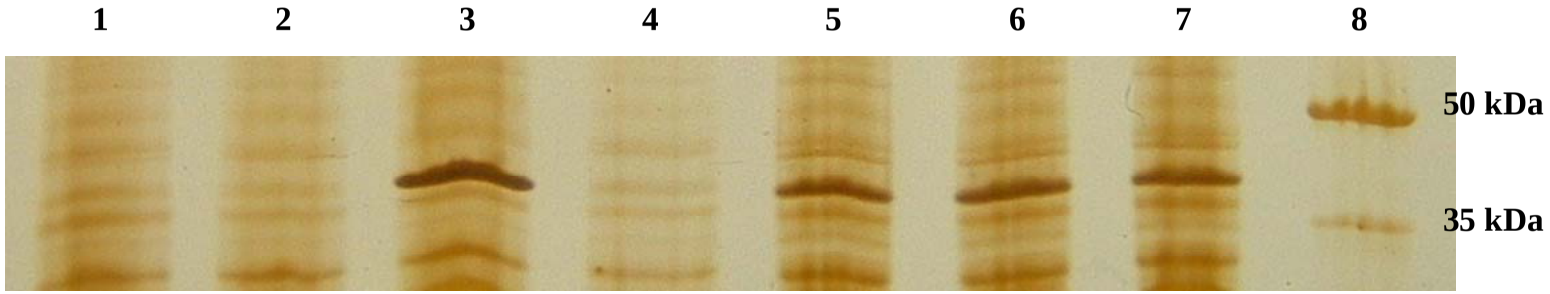


Figure 13. "Mix and match" plasmids. In order to test for orthogonality and to rule out cross-reaction between the MjProRS/MjtRNAPro pair and the endogenous *E. coli* translational machinery, plasmids were constructed for testing orthogonality in a "mix-and-match" experiment. A series of plasmids were constructed that harbored each of the components of the orthogonal pair and of the *E. coli* pair



1. pHC8 (**elastin-CCC**) + pHEC1 (no pair)

2. pHC8 + pHC10 (MjProRS only)

3. pHC8 + pHC15 (Mj pair)

4. pHC8 + pHC17 (MjProRS + EctRNA)

5. pHC8 + pHC19 (EcProRS)

6. pHC8 + pHC20 (EcProRS + EctRNA)

7. pHC8 + pHC21 (EcProRS + MjtRNA)

8. Perfect protein marker

Figure 14. SDS-PAGE analysis with silver staining indicating the results of the "mix and match" experiment. Protein expression was evident in the presence of the Mj pair (Lane 3) and in all cases where the EcProRS was expressed (Lanes 5, 6, 7). Neither the MjProRS nor the $MjtRNA^{Pro}$ alone enabled protein expression (Lane 2, and Fig. 11). The MjProRS was not able to charge the $EctRNA^{Pro}$ with proline suggesting no cross-reaction (Lane 4).

Protein Purification

Following protein expression, purification of the **elastin-CCC** protein was carried out in the UQ27$^{ts,\ ProA}$ strain. SDS-PAGE analysis and MALDI-TOF-MS analysis were conducted to determine the relative purity and mass of the resultant **elastin-CCC** protein. SDS-PAGE analysis indicated that the protein was highly pure (Fig. 15), and MALDI-TOF-MS analysis indicated that the average protein mass obtained from three trials was very similar and within error compared to the theoretical or calculated mass (Fig. 16). The protein yield was 15.6 mg solid protein/ 1 L cell culture. The yield is relatively low but is acceptable due to the high density of proline residues in the target polypeptide sequence as well as expression in the UQ27$^{ts,\ ProA}$ strain where the cells are grown in minimal media at an unnaturally high temperature of 42 °C to inactivate the endogenous EcProRS. Initial growth at 30 ºC may also contribute to a low yield as fewer cells are propagated before reaching log phase growth when induction is initiated. The two plasmid expression system also puts a growth burden on the cells which must achieve resistance to three antibiotics in order to grow.

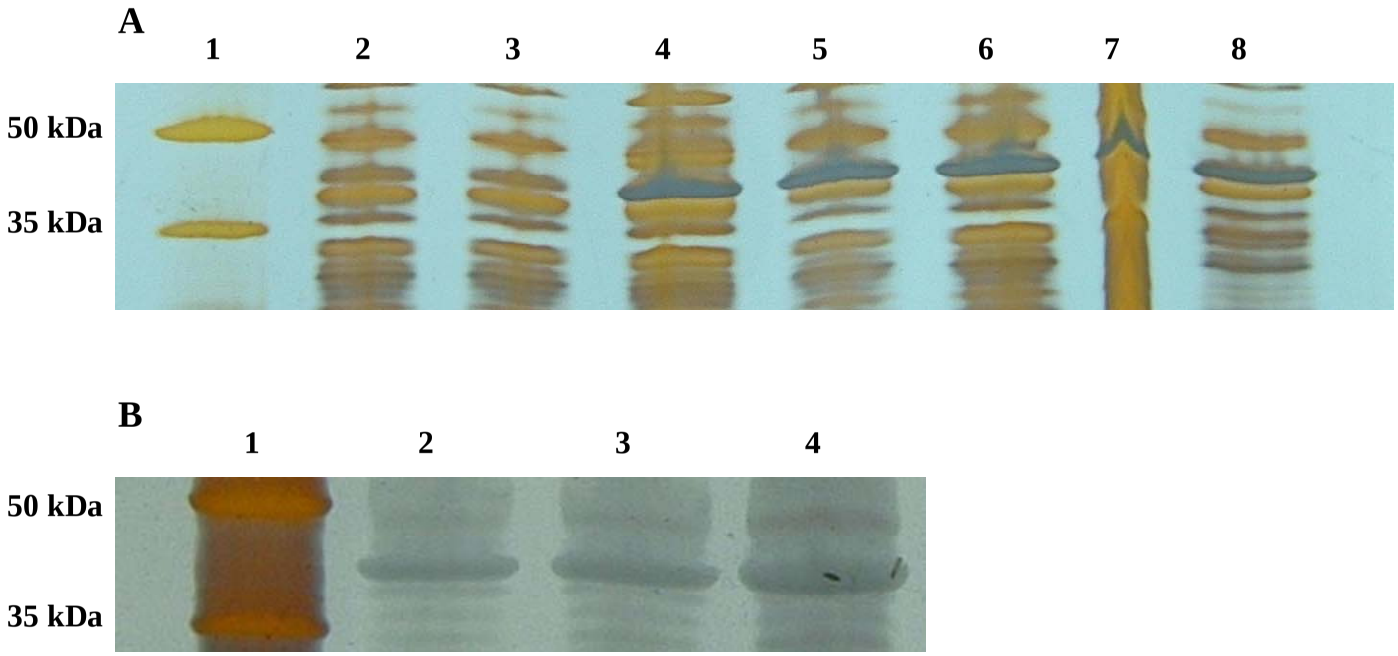


Figure 15. SDS-PAGE analysis with silver staining indicating the results of **elastin-CCC** purification. **Elastin-CCC** was expressed in the UQ27$^{ts,\ ProA}$ strain with IPTG induction. (**A**) Lane 1: Perfect protein marker; Lane 2: Plasmid pHC8 (**elastin-CCC**) only; Lane 3: pHC8 + pHEC1 (no pair); Lane 4: pHC8 + pWK1 (EcProRS); Lane 5: pHC8 + pHC15 (Mj pair); Lane 6: pHC8 + pHC 15; Lane 7 Expression pellet; Lane 8: Expression supernatant (soluble fraction). (**B**) Lane 1: Perfect protein marker; Lane 2: Purified **elastin-CCC**, 0.1 mg/mL; Lane 3: Purified **elastin-CCC**, 0.5 mg/mL; Lane 4: Purified **elastin-CCC**, 1 mg/mL. Slight bands are seen in the purified samples and are characteristic of other purified elastin-mimetic proteins observed previously[29].

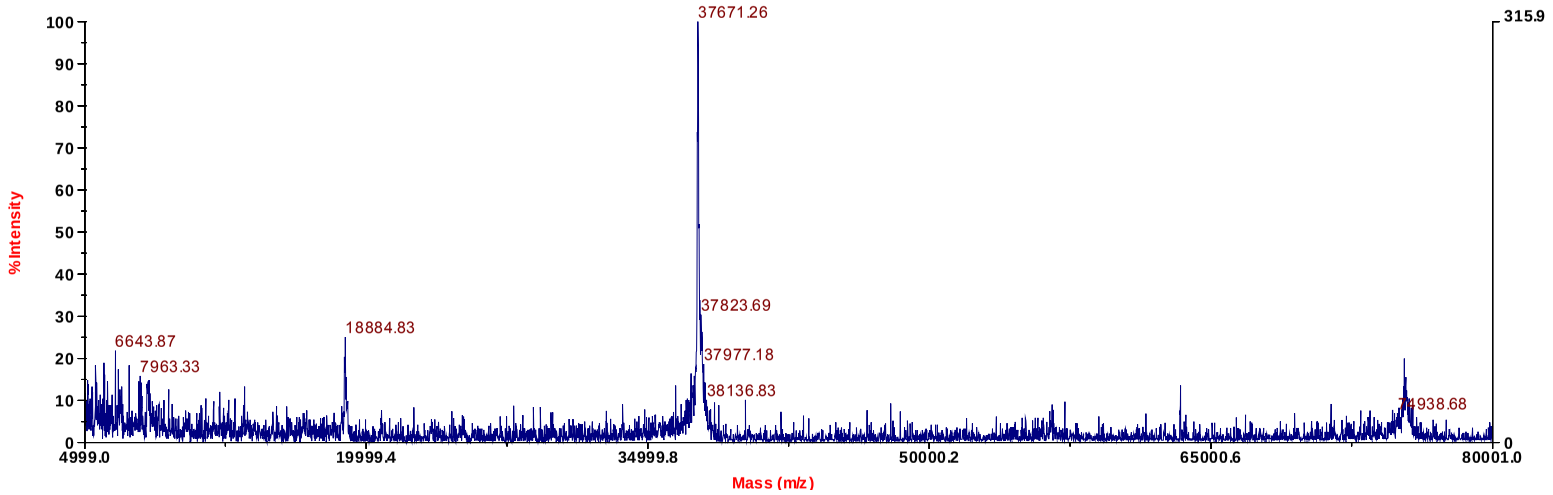


Figure 16. MALDI-TOF-MS mass analysis of purified **elastin-CCC** with L-proline incorporation.

Calculated average mass for **elastin-CCC,** $M[(VPGVG)_4VPGIG]_{17}VPGVGGSD4KGH_{10}$: 37771.10 g/mol

Observed m/z: 37701.17 (average mass calculated from the m/z obtained from three spectra)

Error: 0.19%

Incorporation of Proline Analogues

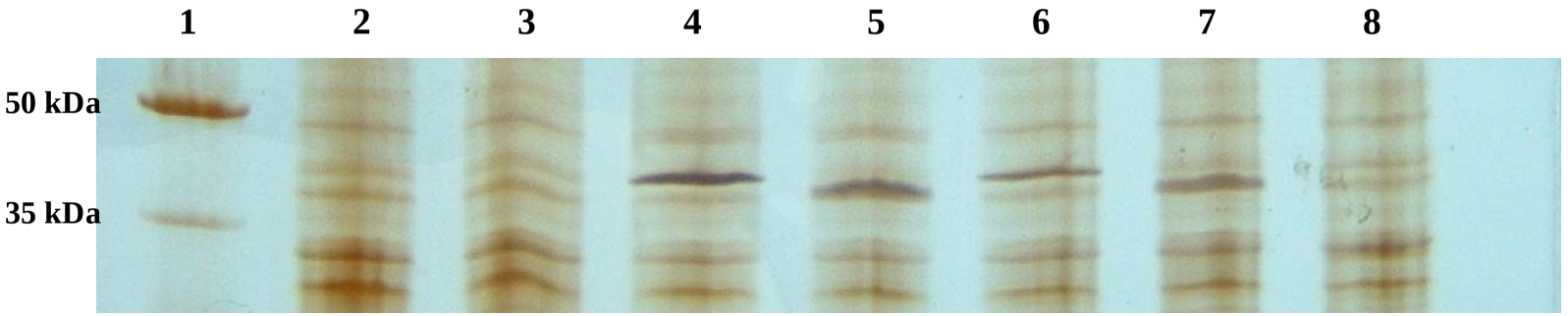
The elastin-mimetic polypeptide, **elastin-CCC**, was employed as a test substrate for assessing the efficacy of multiple site-specific incorporation of proline analogues using a high-level, IPTG-inducible expression system. The repetitive domain of **elastin-CCC** comprises a concatenated series of pentapeptide repeats based on the consensus sequence (Val-Pro-Gly-Val/Ile-Gly) in which proline residues constitute 20 % of the amino acids within the repetitive domain (80 prolines/polypeptide chain). The high density of proline residues and the uniform structural environment of these residues within the polypeptide sequence provide a well-defined context for evaluating the efficacy of analogue incorporation in terms of its effect on protein yield and analogue substitution level. Moreover, using the orthogonal pair for site-specific incorporation allows for the greatest number of substitution events using a site-specific system explored to date. The use of an orthogonal pair to incorporate (*2S,4R*)-4-hydroxyproline at numerous sites throughout a protein polymer, such as in the **elastin-CCC** model polypeptide, is an important step for engineering a system for *in vivo* collagen-mimetic polymer biosynthesis.

The proline analogues, (*2S,4R*)-4-fluoroproline and (*2S,4R*)-4-hydroxyproline were incorporated into the **elastin-CCC** protein using the *M. jannaschii* ProRS/$tRNA^{Pro}$ pair in response to the CCC codons at regular intervals throughout the protein. The test for incorporation of analogues was necessary using the $UQ27^{ts,\ ProA}$ strain to evaluate the host system's utility in further analogue incorporation experiments. It was also necessary to evaluate the ability of the *M. jannaschii* pair to recognize and successfully incorporate proline analogues that exhibit structural differences from the native proline substrate.

Expression of **elastin-CCC** in the pQE-80L vector allowed for high-level expression under tight control of the inducible, phage T5 promoter. A copy of the overproducing lactose repressor, *lacI*^{q}, on the plasmid enabled transcriptional silencing of the target sequence prior to induction. A decahistidine tag was expressed as a C-terminal fusion to the **elastin-CCC** gene as encoded in the adaptor sequence cloned into the plasmid. Structurally similar amino acids have been incorporated into elastin-mimetic proteins in place of a canonical amino acid under selective pressure in which the host bacterium is depleted of the canonical amino acid and supplemented with the analogue. By depleting the growth media of proline, the orthogonal system enables incorporation of two target proline analogues at multiple specific sites throughout the **elastin-CCC** protein as seen in Figures 17 and 18.

The UQ27$^{ts,\ ProA}$ strain was used for incorporation of analogues as other strains, including CAG18515, JM108, and DG99, demonstrated higher levels of background expression without the orthogonal pair present likely due to the activity of the endogenous ProRS enzyme. Conditions for expression of the analogues were optimized as the host strain required a different growth temperature (30 °C) and had a proline auxotrophic phenotype as compared to previous experiments[29]. Minimal media was required for analogue incorporation in order to control the amount of proline in the media at different stages of the expression. Proline depletion with analogue supplementation was sufficient for expression of **elastin-CCC** with the fluorinated analogue, while osmotic shock under high salt conditions was necessary for incorporation of hydroxyproline. With analogue incorporation compared to proline only expression, the electrophoretic mobility of the **elastin-CCC** protein in SDS-PAGE analysis was slightly

altered as described previously with **elastin-1** (Figures 17 and 18)[29]. The *M. jannaschii* pair successfully recognized and enabled incorporation of two target proline analogues. The structurally less-similar analogue, hydroxyproline, required osmotic shock for incorporation as observed previously for incorporation with over-expression of EcProRS[29].



1. Perfect Protein Marker

2. pHC8 + pHEC (no pair), + Proline

3. pHC8 + pHEC (no pair), + Fluoroproline

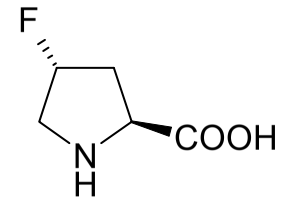
4. pHC8 + pHC15 (Mj pair), + Proline

5. pHC8 + pHC15 (Mj pair), + Fluoroproline

6. pHC8 + pHC15 (Mj pair), + Proline

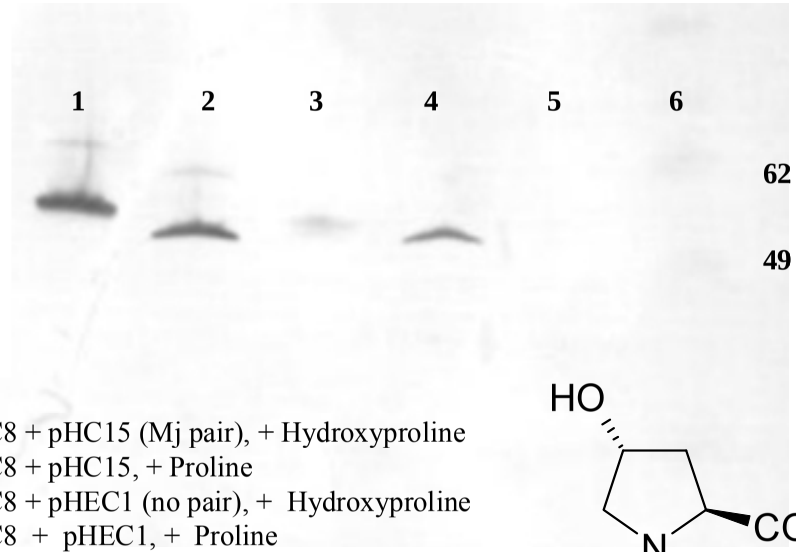
7. pHC8 + pHC15 (Mj pair), + Fluoroproline

8. pHC8 + pHC15, uninduced



(2*S*, 4*R*)-4-fluoroproline

Figure 17. SDS-PAGE analysis with silver staining indicating multi-site incorporation of (2*S*, 4*R*)-4-fluoroproline in **elastin-CCC**. Expression with proline only (Lanes 4 and 6) compared to expression with analogue incorporation (Lanes 5 and 7) indicated a shift in electrophoretic mobility of the **elastin-CCC** protein upon analogue incorporation. Without the orthogonal pair (Lanes 2 and 3), a lack of detectable protein expression was evident with SDS-PAGE analysis. The orthogonal pair enabled protein expression and analogue incorporation.



(2*S*, 4*R*)-4-hydroxyproline

Figure 18. SDS-PAGE analysis with silver staining indicating multi-site incorporation of (2*S*, 4*R*)-4-hydroxyproline in **elastin-CCC**. Expression with proline only (Lane 2) compared to expression with analogue incorporation (Lane 1) indicated a shift in electrophoretic mobility of the **elastin-CCC** protein upon analogue incorporation. Without the orthogonal pair (Lanes 3 and 4), a small amount of detectable protein expression was evident with western blot analysis. The high sensitivity of the technique allowed for visualization of slight background expression without the pair present. This level of expression was much lower than with the pair present (Lanes 1 and 2). The orthogonal pair enabled protein expression and analogue incorporation.

Functional assay for orthogonality

The UQ27$^{ts,\ ProA}$ strain (Genotype: *proS*127(ts), *lacZ*4,lam–, *argG*75)32 was a target for use as a host for the expression of the MjProRS. The spiral assay was used to evaluate the orthogonality of the MjProRS in the E. coli host organism. The UQ27$^{ts,\ ProA}$ cells express a temperature sensitive mutant form of the endogenous prolyl-tRNA synthetase (EcProRS) enzyme. At 30 ºC, the bacteria express an active ProRS enabling incorporation of proline or proline analogues in host cell proteins. At 42 °C, the mutant ProRS is inactive and does not allow for protein synthesis.

At 30 ºC, cells grew normally in each case (Fig. 19, A). However, at 42 °C, the lack of a functional EcProRS prevented cell growth. A recombinant EcProRS, expressed in the cells, allowed for recovery and normal growth, while a recombinant copy of the MjProRS, expressed in the cells, did not allow for growth (Fig. 19, B). This suggests that the *M. jannaschii* synthetase enzyme is orthogonal and that there is no cross-reaction between the MjProRS and the endogenous *E. coli* machinery. The MjProRS was found to be active in the cells as the orthogonal pair shows expression of **elastin-CCC** (Fig. 7)

The orthogonality of the *H. sapiens* synthetase was also evaluated, and expression of the HsProRS did not allow for growth of the cells at 42 °C, which suggests that the HsProRS is either inactive in the cells or that it is functional and orthogonal to the *E. coli* translational machinery. The TtProRS did not allow for growth at 42 °C, which suggests that the TtProRS is either inactive in the cells or that it is functional and orthogonal as well. Expression of the ScProRS in the UQ27 cells at 42 °C allowed for growth, but the growth was diminished compared to the cells which expressed a recombinant copy of the EcProRS.

A. 30 ºC: Permissive

Empty plasmid | *E. coli* synthetase | *M. jannaschii* synthetase

B. 42 ºC: Non- permissive

Empty plasmid | *E. coli* synthetase | *M. jannaschii* synthetase

Figure 19. Results of the functional assay. The UQ27$^{ts, ProA}$ strain (Genotype: *proS*127(ts), *lacZ*4,lam–, *argG*75)[32] was a target for use as a host for the expression of the MjProRS. The bacteria express a temperature sensitive mutant form of the endogenous prolyl-tRNA synthetase (EcProRS) enzyme. At 30 ºC, the bacteria express an active ProRS enabling incorporation of proline or proline analogues in host cell proteins. At 42 °C, the mutant ProRS is inactive and does not allow for protein synthesis. At 30 ºC, cells grow normally in each case (**A**). However, at 42 °C, the lack of a functional EcProRS prevents growth of cells (**B**). A recombinant EcProRS expressed in the cells allows for recovery and normal growth, while a recombinant copy of the MjProRS expressed in the cells does not allow for growth (**B**). This suggests that the synthetase is orthogonal and that there is no cross-reaction between the MjProRS and the endogenous *E. coli* machinery.

Bacterial Host Strain, UQ27 ProRS$^{ts, ProA}$

Gene specific primers were designed to amplify and sequence the EcProRS gene on the genome of the UQ27 ProRS$^{ts, ProA}$ *E. coli* strain. The endogenous, temperature sensitive EcProRS gene had a single nucleotide change (A597T) compared to the wild-type EcProRS of the *E. coli* K12 strain according to sequence alignment and nucleotide BLAST comparison. The single nucleotide change resulted in a mutation in the primary sequence of the EcProRS protein, W375R. This tryptophan to arginine mutation could have caused the temperature sensitive phenotype of the protein due to destabilization of the protein's structure. Upon ClustalW alignment using BioEdit software, neither the TtPRoRS nor the MjProRS proteins have an amino acid which is homologous to W375 in the EcProRS protein (Appendix 1). The archaeal enzymes are missing several short regions that are found in the EcProRS gene. The X-ray crystal structures of the archaeal enzymes are known, but the EcProRS structure has not been solved.

Conclusions

In summary, a bacterial host strain was developed for expression of **elastin-CCC**. A series of plasmids were generated that encoded the HsProRS, EcProRS, ScProRS, and MjProRS genes. In addition, the HstRNA^{Pro}, EctRNA^{Pro}, and MjtRNA^{Pro} genes were generated and cloned into acceptor plasmids. Two orthogonal pairs, one from *H. sapiens* and one from *M. jannaschii*, were constructed using genetic engineering techniques. For each pair, the ProRS and tRNA genes were cloned into a single plasmid for constitutive expression. The human pair appeared to be inactive in the bacterial host as **elastin-CCC** expression was not detected in SDS-PAGE analysis. However, the archaeal pair from *M. jannaschii* was useful in an orthogonal translational system in *E. coli* for the expression

of **elastin-CCC** and for incorporation of the proline analogues, fluoroproline and hydroxyproline. The pair appeared to function independently from the *E. coli* translational machinery and no cross-reaction between the archaeal pair and the endogenous *E. coli* system was observed in *in vivo* expressions ("mix and match") and in a functional assay.

The UQ27$^{ts,\ ProA}$ *E. coli* strain was necessary for achieving low background levels of expression with the orthogonal pair as the endogenous ProRS in the host strain was temperature sensitive and was inactivated with incubation of the cells at 42 °C. A tryptophan to arginine mutation was observed upon sequencing of the endogenous UQ27 ProRS gene that could have caused the temperature sensitive phenotype observed. The elastin derivative, **elastin-CCC,** was purified to homogeneity from the endogenous proteins of the bacterial host using immobilized metal affinity chromatography, and MALDI-TOF-MS analysis was conducted to confirm the size of the purified **elastin-CCC** protein.

References

(1) Wang, L.; Schultz, P. G. *Angewandte Chemie-International Edition* **2005**, *44*, 34-66.

(2) Xie, J. M.; Schultz, P. G. *Nature Reviews Molecular Cell Biology* **2006**, *7*, 775-782.

(3) Ryu, Y. H.; Schultz, P. G. *Nature Methods* **2006**, *3*, 263-265.

(4) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

(5) Liu, W. S.; Brock, A.; Chen, S.; Chen, S. B.; Schultz, P. G. *Nature Methods* **2007**, *4*, 239-244.

(6) Chen, S.; Schultz, P. G.; Brock, A. *Journal of Molecular Biology* **2007**, *371*, 112-122.

(7) Olsen, D.; Yang, C. L.; Bodo, M.; Chang, R.; Leigh, S.; Baez, J.; Carmichael, D.; Perala, M.; Hamalainen, E. R.; Jarvinen, M.; Polarek, J. *Advanced Drug Delivery Reviews* **2003**, *55*, 1547-1567.

(8) Noren, C. J.; Anthonycahill, S. J.; Suich, D. J.; Noren, K. A.; Griffith, M. C.; Schultz, P. G. *Nucleic Acids Research* **1990**, *18*, 83-88.

(9) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537-539.

(10) Mendel, D.; Cornish, V. W.; Schultz, P. G. *Annual Review of Biophysics and Biomolecular Structure* **1995**, *24*, 435-462.

(11) Connor, R. E.; Tirrell, D. A. *Polymer Reviews* **2007**, *47*, 9-28.

(12) Xie, J. M.; Schultz, P. G. *Current Opinion in Chemical Biology* **2005**, *9*, 548-554.

(13) Beuning, P. J.; Musier-Forsyth, K. *Journal of Biological Chemistry* **2001**, *276*, 30779-30785.

(14) Burke, B.; Lipman, R. S. A.; Shiba, K.; Musier-Forsyth, K.; Hou, Y. M. *Journal of Biological Chemistry* **2001**, *276*, 20286-20291.

(15) Wang, L.; Schultz, P. G. *Chemistry & Biology* **2001**, *8*, 883-890.

(16) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 9785-9789.

(17) Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G. *Journal of the American Chemical Society* **2003**, *125*, 935-939.

(18) Zhang, Y.; Wang, L.; Schultz, P. G.; Wilson, I. A. *Protein Science* **2005**, *14*, 1340-1349.

(19) Stehlin, C.; Burke, B.; Yang, F.; Liu, H. J.; Shiba, K.; Musier-Forsyth, K. *Biochemistry* **1998**, *37*, 8605-8613.

(20) Anderson, J. C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 9598-9608.

(21) Magliery, T. J.; Anderson, J. C.; Schultz, P. G. *Journal of Molecular Biology* **2001**, *307*, 755-769.

(22) Santoro, S. W.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 4185-4190.

(23) Santoro, S.; Schultz, P. *Directed evolution of the substrate specificities of a site-specific recombinase and an aminoacyl-tRNA synthetase using fluorescence-activated cell sorting (FACS). Methods in Molecular Biology*

2003; Vol. 230.

(24) Kwon, I.; Kirshenbaum, K.; Tirrell, D. A. *Journal of the American Chemical Society* **2003**, *125*, 7512-7513.

(25) Kane, J. F. *Current Opinion in Biotechnology* **1995**, *6*, 494-500.

(26) Qian, Q.; Bjork, G. R. *Journal of Molecular Biology* **1997**, *273*, 978-992.

(27) Chen, P.; Qian, Q.; Zhang, S. P.; Isaksson, L. A.; Bjork, G. R. *Journal of Molecular Biology* **2002**, *317*, 481-492.

(28) McClain, W. H.; Gabriel, K. *Journal of Molecular Biology* **2001**, *310*, 537-542.

(29) Kim, W. Y.; George, A.; Evans, M.; Conticello, V. P. *Chembiochem* **2004**, *5*, 928-936.

(30) van Hest, J. C. M.; Tirrell, D. A. *Chemical Communications* **2001**, 1897-1904.

(31) Singer, M.; Baker, T. A.; Schnitzler, G.; Deischel, S. M.; Goel, M.; Dove, W.; Jaacks, K. J.; Grossman, A. D.; Erickson, J. W.; Gross, C. A. *Microbiological Reviews* **1989**, *53*, 1-24.

(32) Bohman, K.; Isaksson, L. A. *Molecular & General Genetics* **1980**, *177*, 603-605.

(33) Gabriel, K.; McClain, W. H. *Journal of Molecular Biology* **2001**, *310*, 543-548.

(34) Studier, F. W. *Protein Expression and Purification* **2005**, *41*, 207-234.

(35) Sambrook, J.; Russell, D. W. *Molecular Cloning: A laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

Chapter 4

Development of a Fluorescence-Activated Cell Sorting (FACS)-Based Screening System for use in Altering Substrate Specificity in the *M. jannaschii* Prolyl-tRNA Synthetase Enzyme

Introduction

Directed evolution and FACS-based screening methodologies

Several approaches are available for engineering altered substrate specificity in enzymes[1,2]. Directed evolution techniques have shown recent promise in engineering modified enzymes for various applications[3-5]. Aminoacyl-tRNA synthetase enzymes have been targeted in directed evolution studies and they have been successfully altered for use in orthogonal pairs for incorporation of amino acid analogues in bacterial systems[4,6,7]. For example, Peter Schultz and coworkers optimized a useful method for the evolution of aminoacyl-tRNA synthetase pairs[8]. The group was able to selectively incorporate a series of tyrosine analogues into proteins in *E. coli* after generation of *M. jannaschii* tyrosyl-tRNA synthetase (MjYRS) mutant enzymes specific for the amino acid analogues. The mutant enzyme variants were used with a cognate tRNA as an orthogonal pair in the host for site-specific introduction of *p*-isopropyl-phenylalanine, *p*-amino-phenylalanine, *p*-carboxyl-phenylalanine, and *O*-allyl-phenylalanine. The side-chain binding pocket of the MjYRS enzyme exhibited flexibility in the accommodation of the different functional groups introduced and had been shown previously to selectively accept substrates other than tyrosine[6]. The strategy for generation of synthetase mutants that showed a preference for tyrosine analogues was based on high-throughput screening of large libraries of synthetase variants using reporter detection with fluorescence-activated cell sorting (FACS)[8].

FACS has been used widely in the fields of medicine and immunology for a wide variety of applications. FACS is often used to monitor the progression of diseases such as HIV/AIDS and cancer. Using FACS, one is able to sort heterogeneous mixtures of

cells into separate populations using a fluorescent reporter produced in or presented on the surface of the cells (Fig. 1)[9]. The majority of research involving FACS has been focused on optimization of the sorting technique with eukaryotic cells. The use of FACS presents a challenge in sorting bacterial cells as the cells are not only a different shape (rod-shaped in the case of *E. coli*) compared to roughly spherical eukaryotic cells, but also the diameter of a bacterial cell (1-2 μm) is much smaller that that of an average eukaryotic cell (10-20 μm)[10]. The Schultz group successfully used FACS for rapidly screening large libraries of synthetase variants using synthetase-dependent suppression of amber stop codons with the fluorescent reporter protein, GFPuv[8]. Amber codons within the T7 RNA polymerase gene were recognized by the orthogonal synthetase/tRNA pairs to enable expression of GFPuv in cells where the synthetase variants were active with the analogue of interest, and bacterial cells expressing GFPuv were sorted using FACS in a positive/negative screening strategy alternating with antibiotic selection (chloramphenicol-resistance reporter). Cells harbored a plasmid with a chloramphenicol resistance gene with an amber stop codon position in the DNA sequence of the gene such that read-though of the gene using the orthogonal pair enabled chloramphenicol resistance. Cells expressing mutant synthetase enzymes that showed activity toward the tyrosine analogues could be selected directly using antibiotic resistance.

Directed evolution of the synthetase variants began with a mutagenesis strategy. The crystal structure of a homologous enzyme from *Bacillus stearothermophilus* was examined to determine the residues that made the closest contacts with the substrate in the synthetase enzyme. Five residues were chosen for site-saturation mutagenesis to afford a total library size of 3.2×10^6 (20^5) possible permutations of the gene. The

mutant libraries were subjected to several rounds of positive/negative screening and antibiotic selection to generate synthetase variants that no longer recognized the natural substrate, tyrosine, and were active toward the tyrosine analogues of interest *in vivo*. The group expressed proteins containing the non-canonical amino acids in *E. coli* on a mg protein/L culture scale[8].

David Tirrell and coworkers also used FACS as a method for screening large libraries of mutant synthetase enzymes to generate novel methionyl-tRNA synthetases that enable residue-specific incorporation of non-canonical amino acids in proteins in a bacterial host[11]. A mutant GFP (GFPrm_AM) protein was generated by site-saturation mutagenesis for use as a reporter. As there were only two Met positions in the GFP variant, five additional positions in the primary sequence were changed to Met in order to generate the GFP used as a reporter for analogue incorporation. Four positions within the EcMetRS were chosen for site-saturation mutagenesis as they make close contacts with the natural substrate based on crystal structure data. Several rounds of positive/negative screening (3 positive, 1 negative) were conducted using FACS to enrich for synthetase variants with a preference for 6,6,6-trifluoronorleucine (Tfn) over the natural methionine substrate. Although the *in vitro* kinetic analysis of a mutant synthetase isolated from the screens indicated a sevenfold lower activation of Tfn compared to Met, cells grown in media depleted of Met were able to utilize and incorporate Tfn in GFP using the orthogonal pair. For strains with the mutant MetRS, the yields of proteins synthesized in Tfn-supplemented media were 20-30 mg protein/L culture[11].

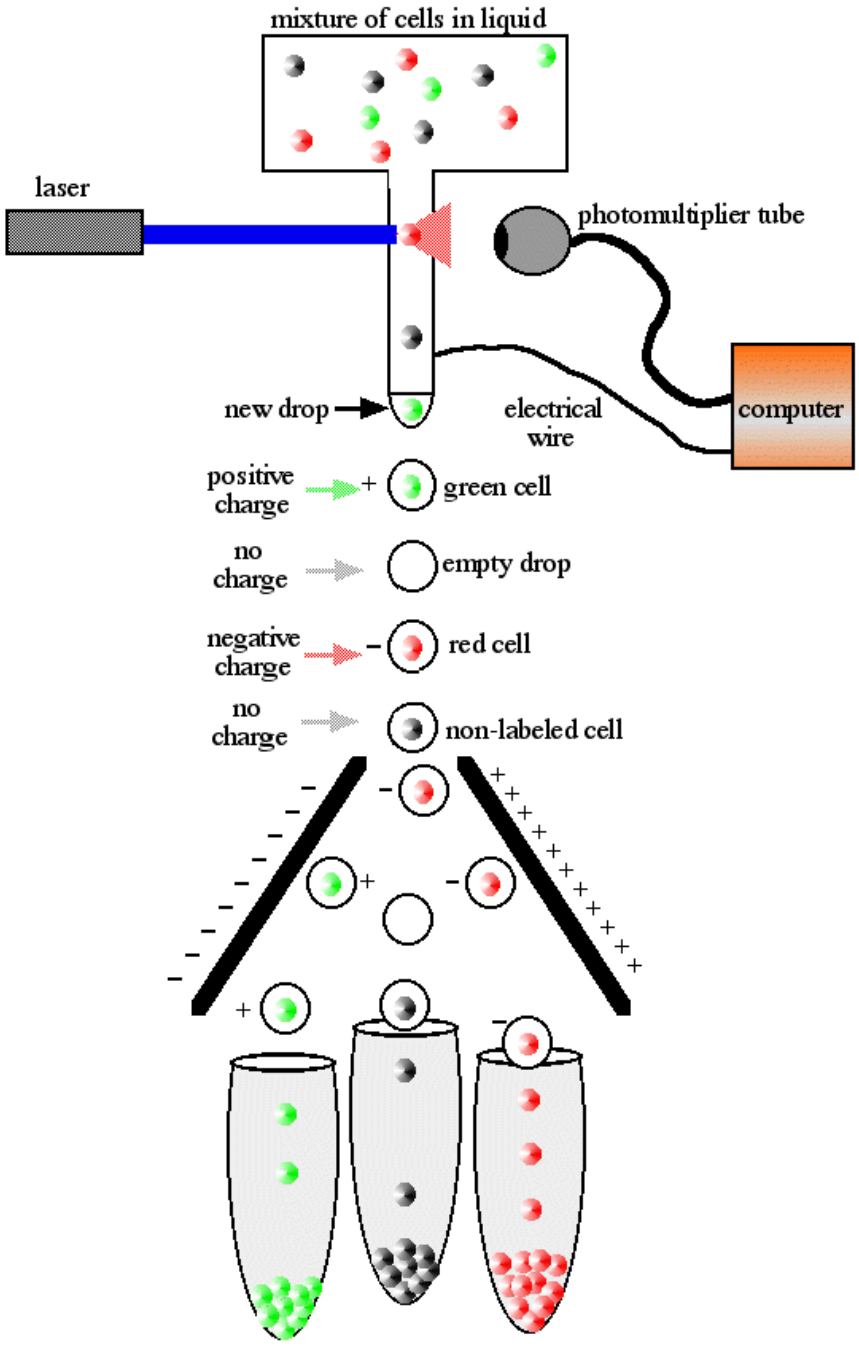


Figure 1. Fluorescence-activated cell sorting (FACS) enables isolation of separate cell populations expressing fluorescent reporter(s) from heterogeneous mixtures of cells. Separate droplets containing cells expressing a fluorescent reporter are detected and assigned a charge. As the charged droplets leave the fluid stream, they are deflected, based on their charge, by the electrical plates into separate collection chambers. The cells remain viable and are then used for further applications. Cells are often sorted in phosphate-buffered saline (PBS) up to 98-99 % purity in each population sorted.

Strategy for screening of prolyl-tRNA synthetase variants

Screening methodologies similar to those developed by the Tirrell and Schultz groups were investigated for engineering site-specificity in orthogonal tRNA/aminoacyl-tRNA synthetase ($tRNA^{Pro}$/ProRS) pairs for multi-site incorporation of proline analogues in *E. coli*. Both of the previously described studies enabled site-specific incorporation of amino acid analogues using an orthogonal pair, but neither study addressed high-level multi-site incorporation of proline analogues in protein-based, polymeric materials. A few sites (1-15) at most were replaced by the analogues of interest. Site-specific incorporation of amino acid analogues in elastin-mimetic polypeptide materials would be an advance in the ability to generate protein polymers with ~20 % non-canonical amino acid content overall compared to incorporation into a few selected sites in catalytic proteins. Generation of a pair with specificity for hydroxyproline over the natural substrate, proline, would be a significant step forward in the development of a route for synthetic collagen biosynthesis in bacteria. Using elastin as a model for collagen biosynthesis would be useful as both elastin and collagen are comprised largely of proline residues. Elastin variants can be readily synthesized using bacterial expression systems, and incorporation of non-canonical amino acids has been demonstrated in elastin-mimetics produced in a bacterial host. With the use of an orthogonal system, tailor-made protein polymer materials can be designed with potentially unique structures and properties.

Rational design, random mutagenesis, and site-saturation mutagenesis of the MjProRS enzyme were the first methods pursued for generation of a mutant MjProRS with specificity for (2*S*,4*R*)-4-hydroxyproline over the natural proline substrate. The

residues, C249 and G251, were chosen as targets for rational design of an enzyme with a preference for hydroxyproline over the natural substrate, proline. The C249 position is homologous to the C443 residue in the EcProRS enzyme that has been implicated as a structurally critical residue in the active site of the EcProRS enzyme as chemical modification of the sulfhydryl side chain was shown to inhibit aminoacylation activity[12]. Mutation of the C443 residue to a glycine residue enabled activity of the mutant EcProRS enzyme with a sterically demanding proline analogue that was not accommodated by the wild type enzyme[13]. Based on crystal structure data, the positions S288 of the TtProRS and C265 of the MtProRS (*M. thermautotrophicus* ProRS) are also homologous to the C249 position in the MjProRS enzyme and are observed to make relatively close contacts between the side chain atoms and the *exo* face of the proline ring. The G251 position was also chosen as a target for rational design as glycine, with a sterically permissive hydrogen side chain, also made close contact with atoms in the proline ring and could potentially be chemically altered to accommodate or induce more favorable binding and activity of the enzyme with hydroxyproline.

Random mutagenesis of enzymes is often based on error-prone PCR amplification strategies, while site-saturation mutagenesis is achieved by using primers for amplification of a gene where selectively chosen amino acid residues are mutated to all of the 20 possible amino acids to generate a library of enzymes. Random mutagenesis has been used in a variety of applications to alter the activity of enzymes[1]. However, the residues that are altered as a result of the procedure might not have a great impact on substrate specificity or activity of the enzyme and may lie outside of the active site[14]. An essentially random technique allows for alteration of amino acid residues outside the

active site that may play an important role in binding of the substrate or activity of the enzyme in the context of the 3D structure[14]. Frances Arnold, in particular, has found success in targeting regions outside the active site for mutagenesis and directed evolution techniques used for altering enzyme activity[15].

As an alternate to random mutagenesis, site-saturation mutagenesis was also pursued as a more targeted method for altering the substrate specificity of the MjProRS enzyme. Based on examination of the crystal structure of the homologous ProRS enzyme from *T. thermophilus* (PDB code: 1H4T), four residues within the active site of the MjProRS enzyme were chosen for site-saturation mutagenesis (E103, F150, E152, and C249) (Fig. 2). Each of the four residues chosen as targets for altering substrate specificity of the enzyme made close contacts (1.5-3.5 Å) with the bound L-proline substrate in the TtProRS homolog. Although the crystal structure of the MjProRS enzyme is known (PDB code: 1NJ8), the structure does not include a bound substrate. The use of the homologous structure from *T. thermophilus* was valuable in determination of the residues chosen for mutagenesis in the MjProRS enzyme. Using crystal structure data does have some limitations as the structure information is a static view of an enzyme but similar approaches have been used previously[6,16].

The *E. coli* prolyl-tRNA synthetase (ProRS) enzyme exhibits a significant preference for the natural substrate, proline, compared to the imino acid analogue, hydroxyproline. The specificity constants (k_{cat}/K_m) for activation of hydroxyproline and proline by the *E. coli* ProRS (EcProRS) enzyme are 0.007 $s^{-1} \cdot mM^{-1}$ and 450 $s^{-1} \cdot mM^{-1}$, respectively, indicating the EcProRS activates hydroxyproline ~5 orders of magnitude less efficiently than proline *in vitro*[17]. Based on the activity of the EcProRS enzyme,

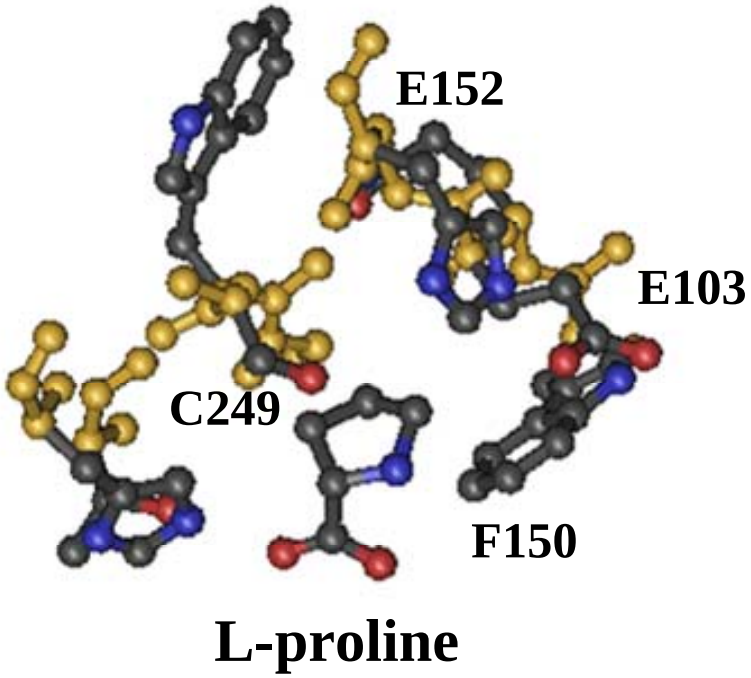


Figure 2. Structure of the active site of the MjProRS enzyme (PDB code: 1NJ8) with residues chosen for site-saturation mutagenesis. Each of the four residues chosen as targets for altering substrate specificity of the enzyme make close contacts (1.5-3.5 Å) with the bound L-proline substrate based on the TtProRS enzyme homolog.

one may assume that using a system in which proline is incorporated into proteins by the EcProRS and hydroxyproline is incorporated by an orthogonal pair (including a synthetase variant with a higher k_{cat}/K_m for hydroxyproline compared to proline) would be an efficient method for simultaneous incorporation of proline and hydroxyproline in protein materials as the affinity of the EcProRS for hydroxyproline is very low. Also, the k_{cat}/K_m values for the MjProRS enzyme were determined by an *in vitro*, ATP hydrolysis assay. The k_{cat}/K_m for the enzyme incubated with L-proline was determined at 60 ºC to be 12,000 $s^{-1} \cdot mM^{-1}$ with a $k_{cat} = 210 \pm 44$ and a $K_m = 0.012 \pm 0.0043$[18]. The kinetic data

collected for the enzyme demonstrate a weaker model for its function as data were collected at 60 ºC compared to *in vivo* activity in an *E. coli* host at 37 ºC, and the k_{cat}/K_m for the *M. jannaschii* enzyme in the presence of hydroxyproline has not been determined.

FACS-based screening approaches were shown to be useful for high-throughput screening of large libraries of synthetase mutants[16], so the technique was chosen for screening the MjProRS mutant libraries. GFPuv was chosen first as a reporter in FACS-based screening of the mutant MjProRS synthetase enzyme libraries as the reporter protein was useful as described previously[8]. GFPuv differs from wild-type GFP as the primary sequence of GFP was altered through mutagenesis to create a GFP variant that was optimized for excitation at UV wavelengths (360-400 nm)[19]. The GFPuv protein contains 10 proline residues that are located on the periphery of the β-barrel structure in regions characterized by random coil structure. Each of the codons in the proline family box are represented in the GFPuv structure (5 prolines are encoded by CCA, 2 prolines are encoded by CCC, 2 prolines are encoded by CCT, and 1 proline is encoded by CCG). It was necessary to convert the CCC codons in GFPuv to CCG codons in order to develop a FACS-based screen for mutant prolyl-tRNA synthetases. The initial design for the screening system involved construction of a fusion protein between **elastin-CCC** and GFPuv. Read-through of the multiple CCC codons in the sequence would be carried out by the orthogonal pair, and GFPuv fused to the C-terminus of the **elastin-CCC** protein would serve as a fluorescent reporter that could be assayed using FACS. The fluorescent reporter would be an indicator of activity of a mutant synthetase with a preference for a proline analogue of interest after successful translation of the entire fusion protein.

Expression of GFPuv, as detected by FACS, would show which cells were expressing an orthogonal pair that was active in the presence of hydroxyproline.

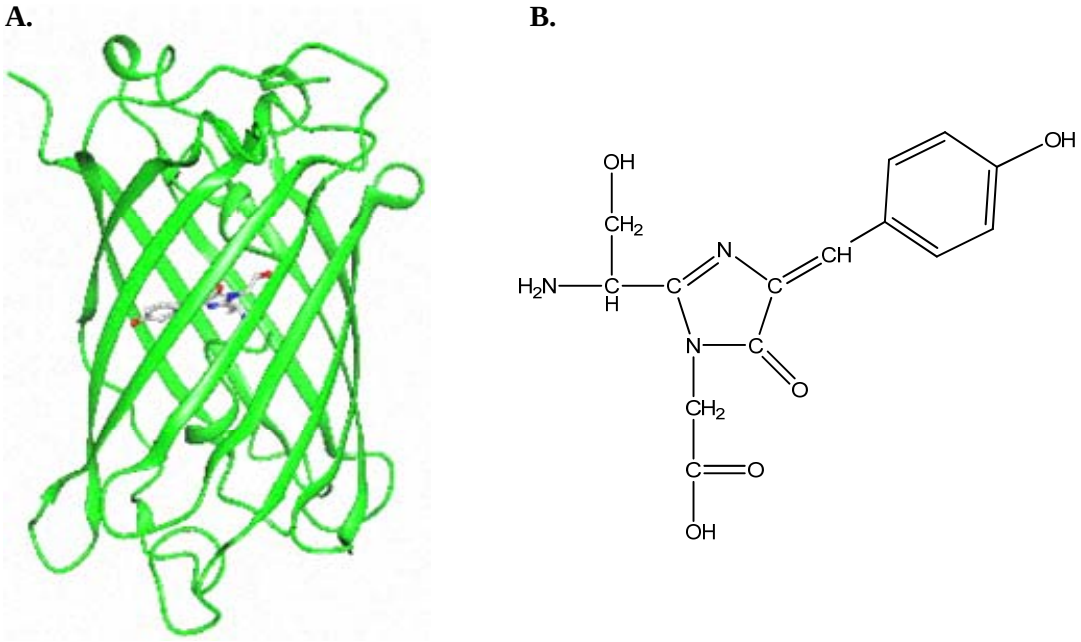
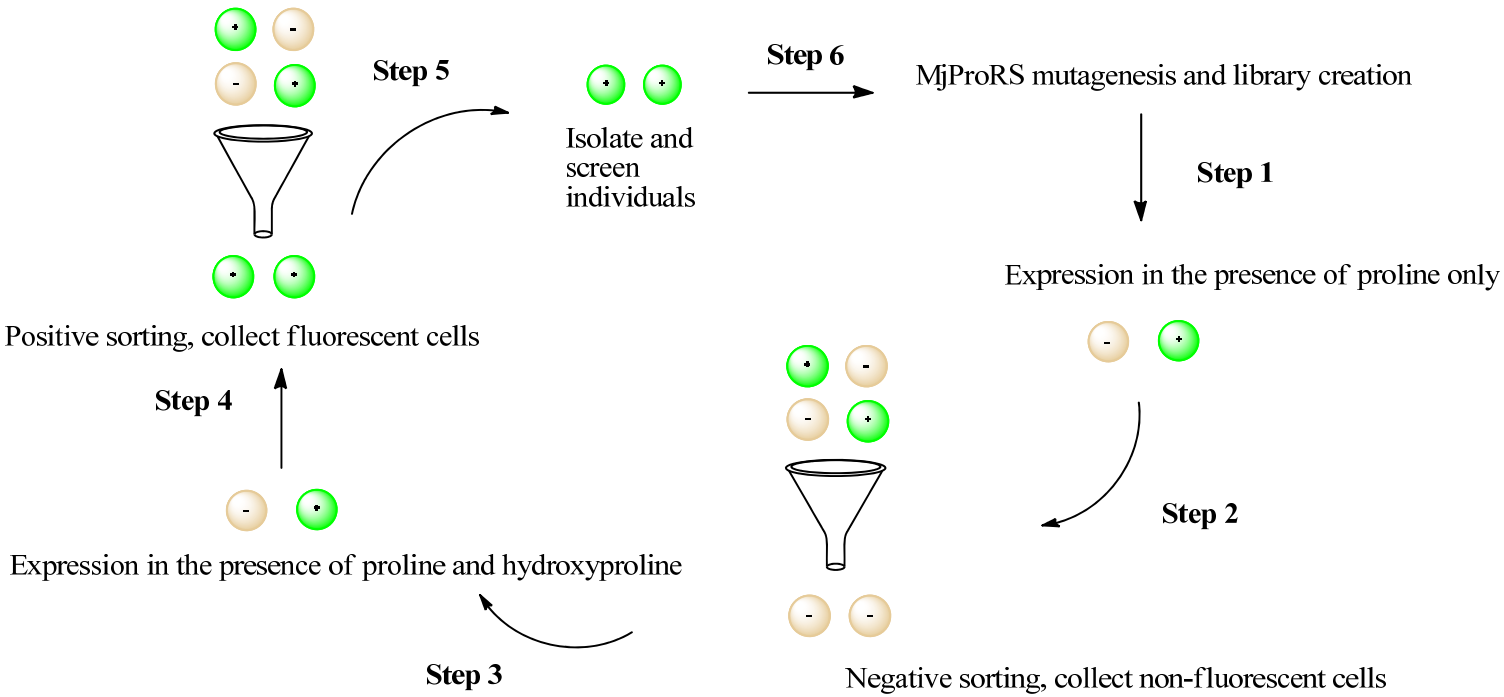


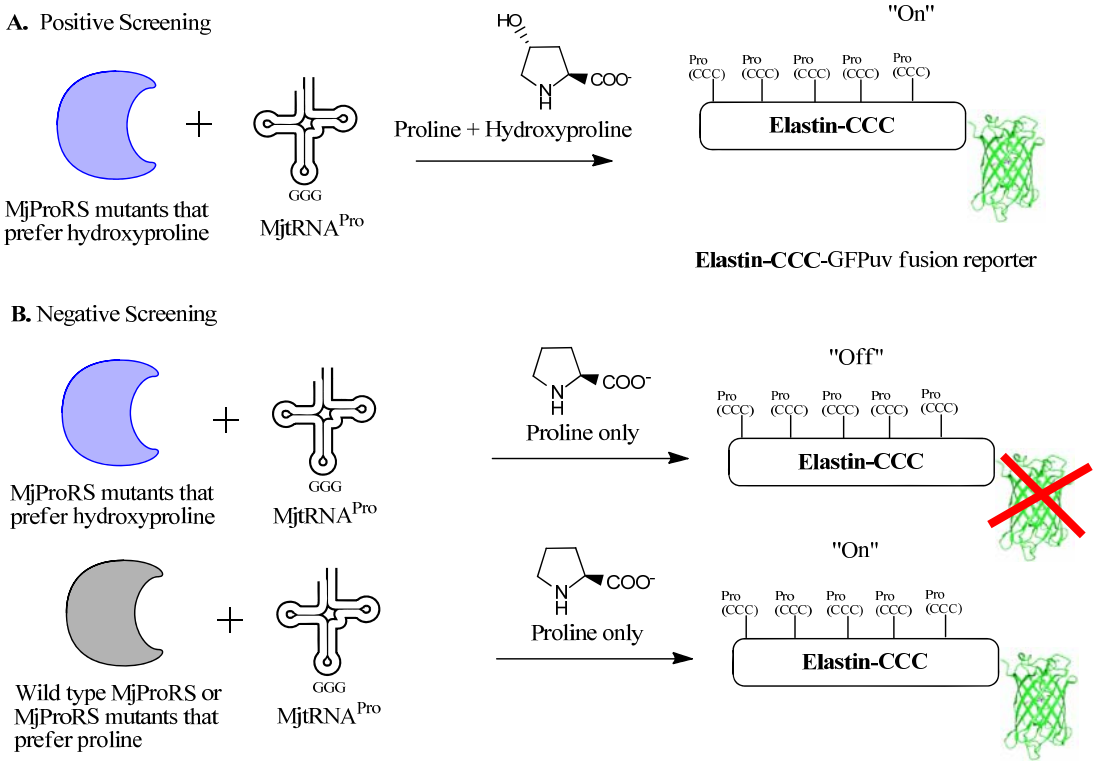
Figure 3. A. Structure of GFP based on the crystal structure (PDB code: 1GFL). **B.** Structure of the chromophore within the GFP protein. The chromophore is formed between Ser65, Tyr66, and Gly67 in the primary sequence of the folded GFP protein.

In the screen, (Schemes 1 and 2) the fusion **elastin-CCC**-GFPuv protein would be expressed in cells also containing a vector for expression of the wild type or mutant orthogonal pair. Cells that were fluorescent in the presence of a mixture of proline and hydroxyproline could contain a pair that accepted proline or hydroxyproline. Cells would be sorted using FACS for the brightest clones. Those same cells would be collected after sorting, and viable cells would be recovered and grown overnight. The following day,

cells would be subjected to expression in the presence of proline only. FACS analysis and sorting following the expression would be used to collect cells that were non-fluorescent in the presence of proline only. The resultant population of cells should show fluorescence with hydroxyproline but not proline and be enriched with synthetase variants that preferred hydroxyproline over proline if the screen and library mutagenesis were effective. The orthogonal pairs within those cells would show a preference for hydroxyproline over the wild-type substrate, proline, and fluorescence of the reporter would be highest in the presence of hydroxyproline compared to proline.



Scheme 1. Positive/Negative Screening Strategy for Isolation of MjProRS Variants with Altered Substrate Specificity. **Step 1.** Create a library of MjProRS mutants. Co-transform UQ27$^{ts, ProA}$ cells with plasmids harboring MjProRS/MjtRNA^{Pro} pairs (library of MjProRS variants) and a plasmid encoding the **elastin-CCC**-GFPuv fusion reporter gene. **Step 2.** Perform an expression at 37 ºC where protein production is induced in the presence of proline only. Cells that are fluorescent contain a MjProRS variant that can charge the MjtRNA with proline and enable translation of the reporter gene. Cells that are non-fluorescent contain a MjProRS variant that is either inactive or is unable to accept proline. **Step 3.** Sort cells using FACS to collect the cells that are non-fluorescent. Repeat sorting as necessary. **Step 4.** Perform an expression at 37 ºC where protein production is induced in the presence of a mixture of proline and hydroxyproline. Cells that are fluorescent contain a mutant MjProRS that can charge the MjtRNA with hydroxyproline and enable translation of the reporter gene. The same cells as seen from Step 2 contain a MjProRS variant that is inactive toward proline. Cells that are non-fluorescent contain an inactive MjProRS variant. **Step 5.** Sort cells using FACS to collect the cells that are fluorescent. Repeat sorting as necessary. **Step 6.** Screen individuals for evidence of a preference for hydroxyproline. Perform further mutagenesis as necessary.



Scheme 2. Rationale behind the positive/negative screening strategy. **A.** In the case of positive screening with the **elastin-CCC**-GFPuv fusion reporter, MjProRS mutants would be co-expressed with the MjtRNA^{Pro} and fusion reporter protein in the presence of proline and hydroxyproline. For MjProRS mutants that prefer hydroxyproline, translational read-through and full-length reporter protein expression would occur due to decoding of the CCC codons in the **elastin-CCC** gene. Cells would appear to be brightly fluorescent with GFPuv production. However, with negative screening in the presence of proline only, (**B.**) MjProRS mutants that prefer hydroxyproline would give rise to non-fluorescent cells as the mutant orthogonal pair could not accept proline and enable translational read-through of the reporter. Only wild type MjProRS or MjProRS mutants that prefer proline would give rise to fluorescent cells. This rationale was used to develop the FACS-based screening strategy in Scheme 1.

Chapters 2 and 3 provide detail concerning the construction and evaluation of the orthogonal system, while this chapter (Chapter 4) gives further detail concerning the construction of mutant synthetase enzymes and libraries in addition to details about the reporter systems developed in FACS-based screening of MjProRS variants for altered substrate specificity toward (2*S*,4*R*)-4-hydroxyproline. Generation of a pair with specificity for hydroxyproline over the natural substrate, proline, would be a significant step forward in the development of a route for synthetic collagen biosynthesis in bacteria. The development of a screening system for generating a novel, orthogonal tRNA/aaRS pair alone is a significant advance in the fields of directed evolution and biomaterials engineering. The screening system and pair would be highly valuable tools in the molecular toolbox available for chemists and biomedical engineers in the development and production of novel protein-based biomaterials.

Materials

All chemical reagents were purchased from either Fisher Scientific, Inc. (Pittsburgh, PA) or Sigma Chemical Co. (St. Louis, MO) unless otherwise noted. Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases, T4 DNA ligase, and T4 kinase were purchased from New England Biolabs, Inc. (Beverly, MA), shrimp alkaline phosphatase was obtained from Roche Applied Science (Indianapolis, IN), and Platinum *Pfx* DNA polymerase was obtained from Invitrogen Corp. (Carlsbad, CA). Plasmid pPROTetE.133 was obtained from BD Biosciences, Inc. (Palo Alto, CA). Plasmid pPROLarA.231 was obtained from Professor Rik Myers of the University of Miami. Plasmid pCS-364, which encodes the native *E. coli* prolyl-tRNA synthetase as an *N*-

terminal hexahistidine fusion in plasmid pQE-30 (QIAGEN, Inc.), was a gift from Professor Karin-Musier Forsyth of the University of Minnesota. *E. coli* strain DG99 was purchased from American Type Culture Collection (ATCC# 47041) and strains CAG18515[20] and UQ27[21] were obtained from the *E. coli* Genetic Stock Center at Yale University. The UQ27 strain was modified to produce a proline auxotrophic phenotype as described previously in Chapter 3. The *E. coli* strain, Top10F´ was obtained from Invitrogen Corp. (Carlsbad, CA. The pSU81 plasmid was a gift from Professor William McClain at the University of Wisconsin[22]. The plasmid pRAM2 was employed as a source of the gene encoding the **elastin-1** sequence[13]. Double digestion of pRAM2 with *Nco* I and *Bam*H I afforded a duplex DNA cassette of approximately 1300 bp, which was inserted into the compatible *Nco* I/*Bgl* II sites of plasmid pQE-60 (QIAGEN, Inc., Valencia, CA) to generate plasmid pAG1. An *Eco*R I/*Hin*D III cassette derived from pAG1 was excised, isolated, and cloned into the compatible sites of plasmid pQE-80L (QIAGEN, Inc., Valencia, CA) to afford pAG2.

Plasmid constructs encoding variants of the *E. coli* prolyl-tRNA synthetase were derived from plasmid pCS-364 (a gift from Professor Karin-Musier Forsyth of the University of Minnesota)[12], and plasmid pWK1 encoding the EcProRS gene was generated by Dr. Wookhyun Kim in the Conticello lab at Emory University[13]. The plasmid, pBAD/HisA, was purchased from Invitrogen Corp. (Carlsbad, CA). The pGFPuv plasmid, containing the gene for GFPuv, and the pDsRed-Express plasmid were purchased from Clontech, Inc. (Mountain View, CA). The superfolder GFP gene was obtained on a pET expression vector as a gift from the lab of Geoffrey Waldo at Los Alamos National Laboratory[23].

Synthetic oligonucleotides were purchased from either Sigma-Genosys, Inc (The Woodlands, TX) or Integrated DNA Technologies (Coralville, IA) and were used as received. TALON® metal affinity resin was purchased from Clontech, Inc. (Mountain View, CA). NMM medium was prepared as described previously[13], with the exception that proline was not added to the medium prior to cell culture. Similarly, SMM media was prepared according to the protocol by Studier[24]. Procedures for the manipulation of DNA, the transformation of competent cells, and the growth and induction of bacterial cultures were adapted from the published literature[25] or instructions supplied by manufacturers. Reagents for the manipulation of bacteria and DNA were sterilized by either autoclave or passage through a 0.2 μm filter. Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. Site-directed mutagenesis was performed using Stratagene's (La Jolla, CA) Quick-Change mutagenesis technique from gene-specific oligonucleotide primers. Random mutagenesis was conducted using Stratagene's GeneMorph II Random Mutagenesis Kit, and site-saturation mutagenesis was carried out using Stratagene's QuikChange Multi Site-Directed Mutagenesis Kit. Automated DNA sequencing was performed at the Emory University Core DNA Sequencing Facility on a Perkin-Elmer ABI Prism model 377 DNA sequencer.

Protein electrophoresis was performed using 10-15 % gradient discontinuous pre-cast sodium dodecyl sulfate (SDS) polyacrylamide gels on a PhastSystem from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ). The Perfect Protein Marker was used as a protein standard for SDS-PAGE and western blot analysis and was purchased from Novagen, Inc. (Madison, WI). Silver staining, used for protein visualization, was carried out using a silver staining kit purchased from GE Healthcare Bio-Sciences Corp.

(Piscataway, NJ). Western blotting was carried out using the PhastSystem (GE Healthcare). The primary antibody used for binding his-tagged proteins was the His-tag monoclonal antibody or the anti-GFP antibody from Clontech, Inc. (Mountain View, CA). The secondary antibody used was the goat anti-mouse secondary antibody, also from Clontech, Inc. Polyhistidine tagged proteins were visualized by chemiluminescent detection using the 1-step NBT/BCIP reagent mixture from Pierce Biotechnology (Thermo Scientific) (Rockford, IL). A BioRad MJ Mini thermocycler (Hercules, CA) was used for conducting amplification and mutagenesis reactions.

FACS analysis and sorting of GFPuv containing cells were carried out on a FACSVantage, SE model, system manufactured by Becton Dickinson (BD Biosciences, San Jose, CA), with hardware for 5-color, high-speed multi-laser sorting. The instrument has 488 nm primary lasers and 650 nm secondary red lasers and a UV laser split from the 488 primary. The UV laser (λ = 365 nm) was used for excitation of GFPuv with a 510/520 nm emission filter. FACS analysis, of cells expressing superfolder GFP, was carried out using the LSR II system also by Becton Dickinson (BD Biosciences, San Jose, CA). The 488 nm laser was used for excitation of superfolder GFP with 505 nm and 530/30 nm emission filters.

Methods

Construction of the elastin-CCC/GFPuv fusion reporter systems

Using primers designed for site-directed mutagenesis, the CCC proline codon sites (P89 and P211) in the GFPuv gene (in the pGFPuv plasmid) were both converted to CCG sites. A single *Nco* I site was also removed by silent site-directed mutagenesis from the GFPuv coding sequence to facilitate cloning of the gene into the pBAD/HisA vector.

A C-terminal **elastin-CCC**-GFPuv gene was constructed first. Gene-specific primers were designed for amplification of the GFPuv gene from pGFPuv for insertion of restriction cloning sites at the 5′ and 3′ termini of the gene. The primers were designed for in-frame incorporation of the restriction sites (used for cloning of the gene), for removal of the stop codons and decahistidine tag at the 3′ end of the gene, and for incorporation of a double glycine linker to be placed between the end of the **elastin-CCC** gene and the start codon of the GFPuv gene. The 717 bp GFPuv modified (CCC and *Nco* I site removed) gene was double-digested from the GFPuv plasmid using the *Bgl* II and *Hind* III enzymes. The plasmid pHC8, containing the **elastin-CCC** gene, was double-digested with *Bam*H I/*Hind* III restriction endonucleases. The adaptor sequence in which the **elastin-CCC** multimer was originally cloned was designed with a *Bam*H I cleavage site at the 3′ end of the multimer cassette and just upstream of a decahistidine tag sequence so that when the GFPuv gene was cloned in, the polyhistidine tag was removed. GFPuv could not be directly digested with *Bam*H I as there is an internal *Bam*H I site in the gene. Digestion with *Bgl* II and *Bam*H I leaves the same four base sticky end, GATC, so that the GFPuv gene could be ligated into the acceptor plasmid and therefore cloned directly into the expression vector, pHC8, to generate plasmid pHC43. The plasmid was sequenced to confirm the location and sequence of the GFPuv gene.

An N-terminal GFPuv-**elastin-CCC** gene was constructed to use for comparison of expression and fluorescence levels between N-terminal and C-terminal fusion reporters. Gene-specific primers were designed for amplification of the modified GFPuv gene from the pGFPuv plasmid where *Nco* I sites would be incorporated at both the 5′ and 3′ termini of the gene. Since cloning of the **elastin-CCC** cassette was achieved by

cloning into the pQE-80L vector from the pQE-60 vector, the only site available for cloning of an N-terminal fusion GFPuv was the *Nco* I site used for cloning the **elastin-CCC** gene into the adaptor sequence. A glycine linker was also engineered between the GFPuv gene and the start of the **elastin-CCC** gene to allow for increased flexibility and proper folding of the reporter fusion protein. The GFPuv gene was amplified from the pGFPuv plasmid and the DNA was isolated by agarose gel electrophoresis. The acceptor plasmids, pBAD/HisA and pHC8 were both digested separately with *Nco* I and gel isolated. Plasmids were generally dephosphorylated with shrimp alkaline phosphatase before ligation. The modified GFPuv gene was ligated in two separate reactions into the acceptor plasmids and the gene was sequenced to determine the correct orientation for the GFPuv insert in both plasmids to generate plasmids pHC42 (GFPuv-**elastin-CCC** in pBAD/HisA) and pHC44 (GFPuv-**elastin-CCC** in pQE-80L).

Construction of the elastin-CCC-superfolder GFPuv fusion reporter

The superfolder GFP gene was obtained on a pET expression vector as a gift from the lab of Dr. Geoffrey Waldo at Los Alamos National Laboratory. Using primers designed for site-directed mutagenesis, the two CCC proline codon sites in the superfolder GFP gene were both converted to CCG sites producing silent mutations where each site would still encode proline but the presence of the CCC codons would not interfere with expression of the reporter. The 741 bp modified superfolder GFP (CCC sites removed, with a C-terminal hexahistidine tag) gene was amplified from the pET plasmid using primers designed to amplify the gene with incorporation of *Bgl* II and *Hind* III cloning sites. The plasmid pHC8, containing the **elastin-CCC** gene, was double-digested with *Bam*H I/*Hind* III restriction endonucleases. The adaptor sequence in which

the **elastin-CCC** multimer was originally cloned was designed with a *Bam*H I cleavage site at the 3′ end of the multimer cassette and just upstream of a decahistidine tag sequence. Superfolder GFP could not be directly digested with *Bam*H I as there is an internal *Bam*H I site in the gene. Digestion with *Bgl* II and *Bam*H I leaves the same four base sticky end, GATC, so that the superfolder GFP gene could be ligated into the acceptor plasmid and therefore cloned directly into the expression vector, pHC8, to generate plasmid pHC51. The **elastin-CCC** superfolder GFP gene fusion contained the hexahistidine tag originally engineered into the sequence at the 3′ end of the fusion gene. The plasmid was sequenced to confirm the location and sequence of the superfolder gene.

The superfolder GFP gene, alone, was also cloned directly into the pQE-80L vector to be used as a control construct. The superfolder GFP gene was amplified from the pET vector using gene specific primers that enabled incorporation of *Bam*H I and *Hind* III restriction endonuclease sites at the 5′ and 3′ termini of the gene. The superfolder GFP gene was amplified from the plasmid and the DNA was isolated using the QIAGEN PCR purification kit. The insert and acceptor plasmid, pQE-80L, DNA were both *Bam*H I/*Hind* III double-digested and isolated using agarose gel electrophoresis. The superfolder GFP insert DNA was ligated into the pQE-80L plasmid to generate plasmid, pHC52. The plasmid was sequenced to confirm the location and sequence of the superfolder gene.

Cloning and mutagenesis of DsRed

Gene-specific primers were designed for amplification of DsRed with incorporation of *Bam*H I and *Hind* III restriction endonuclease sites at the 5′ and 3′ termini of the gene. The DsRed gene was amplified from the pDsRed-Express plasmid

and the DNA was isolated using the QIAGEN PCR purification kit. The insert and acceptor plasmid, pQE-80L, DNA were both *Bam*H I/*Hind* III double-digested and isolated using agarose gel electrophoresis. The DsRed insert DNA was ligated into the pQE-80L plasmid to generate plasmid, pHC47. The plasmid was sequenced to confirm the location and sequence of the DsRed gene.

Site-saturation mutagenesis of DsRed was carried out to randomize the P63 (proline 63) near the chromophore forming residues. Primers were designed in which the P63 position would be mutated to all possible amino acids upon amplification of the DsRed gene. Randomization of the primers was conducted during synthesis where the position to be randomized (to generate a pool of primers for the mutagenesis reaction) was encoded by either NNK in the forward direction or MNN in the reverse direction (N = A, C, G, or T; K = G or T; M = A or C) to reduce the incorporation of stop codons.

DsRedP63Mut-F

5′-CGCCTGGGACATCCTGTCCNNKCAGTTCCAGTACGGCTCCAAGG-3′

DsRedP63Mut-R

5′-CCTTGGAGCCGTACTGGAACTGMNNGGACAGGATGTCCCAGGCG-3′

Amplification of DsRed was carried out by PCR amplification using the mutagenic primers to generate a pool of DsRed mutant genes randomized at the P63 position. Reactions for mutagenesis were prepared as follows. The final concentrations of each component included 1X P*fx*50 amplification buffer (5 µL, 10X stock), 0.3 mM each dNTP (1.5 µL, 10 mM dNTP mixture), 0.3 µM each primer (1.5 µL, 10 µM stock each), 500 ng template DNA, 1.0-2.0 units P*fx*50 enzyme and water to fill to 50 µL total volume in a thin-walled PCR tube. The solutions were mixed by pipette and pulsed to

bring the solutions to the bottom of the tube. Care was taken not to introduce bubbles in the solutions. The tube(s) were placed in the heat block of a thermocycler and subjected to the following amplification regime; Step 1: 94 ºC for 5 min, Step 2: 94 ºC for 30 s, Step 3: 52 ºC for 30 s, Step 4: 68 ºC for 5 min (1 min/kb), Step 5: Repeat steps 2-4 30 times, Step 6: 68 ºC for 7 min, Step 7: 4 ºC forever.

The DNA was purified following amplification using the QIAGEN PCR purification kit and the DNA was run on an agarose gel to confirm the size of the product generated. The parent plasmid DNA (methylated) was removed by digestion with the *Dpn* I enzyme and then the DNA was transformed directly into *E. coli*. The mutant gene and the acceptor plasmid, pQE-80L, DNA were both *Bam*H I/*Hind* III double-digested and isolated using agarose gel electrophoresis. The DsRed mutant insert DNA pool was ligated into the pQE-80L plasmid, and several colonies containing the plasmid of interest were sequenced to confirm the location and sequence of the DsRed gene. 8 of the colonies isolated on an LB solid plate (containing ampicillin, 100 μg/mL and IPTG, 1 mM) after ligation and transformation that were visibly fluorescent were picked for screening by sequencing.

Cloning and mutagenesis of GFPuv

Site-saturation mutagenesis of GFPuv was carried out to randomize the P89 (proline 89) residue. Primers were designed in which the P89 position would be mutated to all possible amino acids upon amplification of the GFPuv gene. Randomization of the primers was conducted during synthesis where the position to be randomized (to generate a pool of primers for the mutagenesis reaction) was encoded by either NNK in the

forward direction or MNN in the reverse direction (N = A, C, G, or T; K = G or T; M = A or C) to reduce the incorporation of stop codons.

GFPuvP89-F

5′-CATGACTTTTTCAAGAGTGCCATGNNKGAAGGTTATGTACAGGAACGCAC-3′

GFPuvP89-R

5′-GTGCGTTCCTGTACATAACCTTCMNNCATGGCACTCTTGAAAAAGTCATG-3′

Amplification of GFPuv was carried out by PCR amplification using the mutagenic primers to generate a pool of GFPuv mutant genes randomized at the P89 position. Reactions for mutagenesis were prepared as follows. The final concentrations of each component included 1X P*fx*50 amplification buffer (5 µL, 10X stock), 0.3 mM each dNTP (1.5 µL, 10 mM dNTP mixture), 0.3 µM each primer (1.5 µL, 10 µM stock each), 500 ng template DNA, 1.0-2.0 units P*fx*50 enzyme and water to fill to 50 µL total volume in a thin-walled PCR tube. The solutions were mixed by pipette and pulsed to bring the solutions to the bottom of the tube. Care was taken not to introduce bubbles in the solutions. The tube(s) were placed in the heat block of a thermocycler and subjected to the following amplification regime; Step 1: 94 ºC for 5 min, Step 2: 94 ºC for 30 s, Step 3: 52 ºC for 30 s, Step 4: 68 ºC for 5 min (1 min/kb), Step 5: Repeat steps 2-4 30 times, Step 6: 68 ºC for 7 min, Step 7: 4 ºC forever.

The DNA was purified following amplification using the QIAGEN PCR purification kit and the DNA was run on an agarose gel to confirm the size of the fragment generated. The parent plasmid DNA (methylated) was removed by digestion with the *Dpn* I enzyme, and then the DNA was transformed directly into *E. coli*. Eighteen of the colonies, isolated on an LB solid plate (containing ampicillin, 100 µg/mL

and IPTG, 1 mM) after ligation and transformation, were picked for screening by sequencing.

The P89 position was mutated to code for alanine in GFPuv using site-directed mutagenesis. Briefly, primers were designed to alter the codon of interest in the GFPuv gene and the gene was amplified using the mutagenic primers to generate the GFPuv P89A mutant. The parent plasmid DNA (methylated) was removed by digestion with the *Dpn* I enzyme, and then the DNA was transformed directly into *E. coli*. Colonies were picked for screening by fluorescence activity and DNA sequencing.

Rational design: Site-directed mutagenesis of MjProRS

Site-directed mutagenesis of the MjProRS gene was conducted using a modified procedure from Invitrogen. Primers were designed for mutation of the sites of interest by altering the coding sequence at each position. Two sites were targeted for site mutations. The C249 and G251 sites were mutated to generate the following mutant MjProRS genes; single mutant MjProRS C249G, double mutant MjProRS C249G, G251T, double mutant MjProRS C249G, G251A, double mutant MjProRS C249G, G251S, and double mutant MjProRS C249G, G251C. The primers used for mutagenesis are listed below.

MjProRSC250G-F

5′- GATTATGCTTATCAAACAGGCTACGGAATCTCAGATAGGG-3′

MjProRSC250G-R

5′-CCCTATCTGAGATTCCGTAGCCTGTTTGATAAGCATAATC-3′

MjaG252T-F

5′- GATTATGCTTATCAAACAGGCTACACTATCTCAGATAGGGTTATAGC-3′

MjaG252T-R

5′-GCTATAACCCTATCTGAGATAGTGTAGCCTGTTTGATAAGCATAATC-3′

MjaG252A-F

5′- GATTATGCTTATCAAACAGGCTACGCTATCTCAGATAGGGTTATAGC-3′

MjaG252A-R

5′-GCTATAACCCTATCTGAGATAGCGTAGCCTGTTTGATAAGCATAATC-3′

MjaG252S-F

5′- GATTATGCTTATCAAACAGGCTACAGCATCTCAGATAGGGTTATAGC-3′

MjaG252S-R

5′-GCTATAACCCTATCTGAGATGCTGTAGCCTGTTTGATAAGCATAATC-3′

MjaG252C-F

5′- GATTATGCTTATCAAACAGGCTACTGCATCTCAGATAGGGTTATAGC-3′

MjaG252C-R

5′-GCTATAACCCTATCTGAGATGCAGTAGCCTGTTTGATAAGCATAATC-3′

Amplification of MjProRS was carried out by PCR amplification using the mutagenic primers to generate mutant genes altered at one or two positions. Reactions for mutagenesis were prepared as follows. The final concentrations of each component included 1X P*fx*50 amplification buffer (5 µL, 10X stock), 0.3 mM each dNTP (1.5 µL, 10 mM dNTP mixture), 0.3 µM each primer (1.5 µL, 10 µM stock each), 500 ng template DNA, 1.0-2.0 units P*fx*50 enzyme and water to fill to 50 µL total volume in a thin-walled PCR tube. The solutions were mixed by pipette and pulsed to bring the solutions to the bottom of the tube. Care was taken not to introduce bubbles in the solutions. The tube(s) were placed in the heat block of a thermocycler and subjected to the following amplification regime; Step 1: 94 ºC for 5 min, Step 2: 94 ºC for 30 s, Step 3: 52 ºC for

30 s, Step 4: 68 ºC for 5 min (1 min/kb), Step 5: Repeat steps 2-4 30 times, Step 6: 68 ºC for 7 min, Step 7: 4 ºC forever.

The DNA was purified following amplification using the QIAGEN PCR purification kit and the DNA was run on an agarose gel to confirm the size of the fragment generated. The parent plasmid DNA (methylated) was removed by digestion with the *Dpn* I enzyme, and then the DNA was transformed directly into *E. coli*. And cells were plated on a LB solid plate with chloramphenicol (34 μg/mL). For the single mutant, MjProRS C249G, the gene was cut from the template plasmid after mutagenesis with *Kpn* I/*Xba* I and re-ligated into either pHEC1 to generate pHC18 or into pHEC1 containing the MjtRNA gene to generate plasmid pHC16. The double mutant genes were double-digested with *Kpn* I/*Xba* I and cloned into pHEC1 with the MjtRNA cassette to generate plasmids pHC22, pHC23, pHC24, and pHC25.

Library construction: Random mutagenesis of MjProRS

Random mutagenesis of the MjProRS gene was carried out using a modified protocol based on the instructions included in Stratagene's GeneMorph II Random Mutagenesis Kit. Gene-specific primers were designed for amplification of the MjProRS gene in the pHC15 plasmid.

MjProRSRM1-F

5′-GCATGGGTACCATGTTGGAATTTTCAGAATGG-3′

MjProRSRM1-R

5′-GCGATTCTAGATTAGTAGGTTTTAGCTATTGC-3′

Two reactions were prepared in which the frequency of mutagenesis was at a medium level in one and high in the other. Amplification of MjProRS was carried out by

PCR amplification using the mutagenic primers to generate mutant genes altered at random positions due to the activity of the error-prone Mutazyme II DNA polymerase. Reactions for mutagenesis were prepared as follows. The final concentrations of each component included 1X amplification buffer (5 μL, 10X stock), 0.3 mM each dNTP (1.5 μL, 10 mM dNTP mixture), 0.3 μM each primer (1.5 μL, 10 μM stock each), 1060 ng template DNA (Medium) or 530 ng (High), 1.0-2.0 units Mutazyme enzyme and water to fill to 50 μL total volume in a thin-walled PCR tube. The solutions were mixed by pipette and pulsed to bring the solutions to the bottom of the tube. Care was taken not to introduce bubbles in the solutions. The tube(s) were placed in the heat block of a thermocycler and subjected to the following amplification regime; Step 1: 95 ºC for 5 min, Step 2: 95 ºC for 1 min, Step 3: 52 ºC for 1 min, Step 4: 72 ºC for 3 min (1 min/kb), Step 5: Repeat steps 2-4 30 times, Step 6: 72 ºC for 10 min, Step 7: 4 ºC forever.

The DNA was purified following amplification using the QIAGEN PCR purification kit and the DNA was run on an agarose gel to confirm the size of the fragment generated. The parent plasmid DNA (methylated) was removed by digestion with the *Dpn* I enzyme, and the mutant MjPRoRS was cloned back into the acceptor (pHC15) plasmid using the *Kpn* I/*Xba* I restriction sites. Several individuals were picked and the DNA was isolated from each to determine the extent of mutagenesis by screening with DNA sequencing. The random mutant library DNA was purified by butanol precipitation and transformed into the $UQ27^{ts,\ ProA}$ strain. The $UQ27^{ts,\ ProA}$ strain, containing the reporter plasmid (elastin-GFP fusion gene) was made electrocompetent the same day as transformation to ensure high transformation efficiency. Electroporation

was used for transformation of the DNA as this method affords the highest transformation efficiency compared to chemical transformation in bacteria.

Library construction: Site-saturation mutagenesis of MjProRS

Site-saturation mutagenesis of the MjProRS gene was carried out using a modified protocol based on the instructions included in Stratagene's QuikChange Multi Site-Directed Mutagenesis Kit. Primers were designed for amplification of the MjProRS gene in the pHC15 plasmid in which four amino acids of interest would be randomized to potentially each of the 20 different canonical amino acids. Four residues within the active site of the MjProRS enzyme were chosen for site-saturation mutagenesis including E103, F150, E152, and C249. Primers were designed for mutagenesis of each of the four residues.

SSMLMjProRS1-F (Mutation of E103)

5′-GCTTTAAGACCTACTTCANNKACACCAATATACTATATG-3′

SSMLMjProRS2-F (Mutation of F150 and E152)

5′-GGTTAAGAGAGATAATGACANNKAAANNKGCCCACACTGCCCATTC-3′

SSMLMjProRS3-F (Mutation of C249)

5′-GATTATGCTTATCAAACANNKTACGGAATCTCAGATAGGG-3′

Amplification of MjProRS was carried out by PCR amplification using the mutagenic primers. Reactions for mutagenesis were prepared as follows. The concentrations of each component included 1X amplification buffer (2.5 µL, 10X stock), 1 µL, 10 mM dNTP mixture, 1 µL each primer (10 µM stock each), 50 ng template DNA, 1.0-2.0 units enzyme blend and water to fill to 25 µL total volume in a thin-walled PCR tube. The solutions were mixed by pipette and pulsed to bring the solutions to the bottom

of the tubes. Care was taken not to introduce bubbles in the solutions. The tubes were placed in the heat block of a thermocycler and subjected to the following amplification regime; Step 1: 95 ºC for 2 min, Step 2: 95 ºC for 1 min, Step 3: 55 ºC for 1 min, Step 4: 65 ºC for 8 min (1 min/kb), Step 5: Repeat steps 2-4 30 times, Step 6: 65 ºC for 7 min, Step 7: 4 ºC forever. Once the reactions reached Step 7, 2 µL *Dpn* I enzyme were added to the tubes and the reactions were incubated at 37 ºC for 4 h.

The DNA in each reaction was purified by butanol precipitation and Top10F´ high efficiency electrocompetent cells were prepared the same day. The DNA was transformed into the Top10F´ high efficiency electrocompetent cells the same day and all of the cells were plated on large library plates (~200 mL solid LB media supplemented with chloramphenicol (34 µg/mL)). One µL of each transformation suspension was added to 99 µL SOC and this was plated on a small Petri plate in order to evaluate the transformation efficiency. Plates were incubated overnight at 37 °C and the number of colonies was counted the next day. Sixteen clones were selected at random for sequencing to evaluate the diversity of the library. The large library plate was scraped using a sterile spreader after addition of 4 mL LB media. The 4 mL suspension was added to 16 mL LB supplemented with chloramphenicol (34 µg/mL) and grown overnight at 37 ºC with shaking at 225 rpm. Two 5 mL aliquots were taken from the 20 mL total culture and plasmid DNA was isolated from the cells using the QIAGEN spin miniprep kit. The concentrated DNA sample was then used for same-day transformation of the library DNA into the high efficiency electrocompetent UQ27 expression strain already harboring the reporter plasmid. Frozen stocks were prepared by adding 800 µL

cell culture to 200 µL 80 % glycerol solution and storage at -80 ºC. Frozen stocks or the original solution were used to inoculate expression seed cultures.

The pHEC2 plasmid was constructed for the cloning of mutant MjProRS genes in the expression plasmid (modified pPROTet/Lar). The pHEC2 plasmid was derived from the pHEC1 plasmid so that the MjtRNA^{Pro} expression cassette could be cloned into the *Nhe* I/*Pvu* I sites used previously for cloning the tRNA genes into the plasmid. The internal *Pvu* I restriction site was removed from pHEC1 by site-directed mutagenesis using primers designed to create a silent mutation in the plasmid to remove the site of interest.

List of primers used in this study

DsRedP63Mut-F

5′-CGCCTGGGACATCCTGTCCNNKCAGTTCCAGTACGGCTCCAAGG-3′

DsRedP63Mut-R

5′-CCTTGGAGCCGTACTGGAACTGMNNGGACAGGATGTCCCAGGCG-3′

GFPuvP89-F

5′-CATGACTTTTTCAAGAGTGCCATGNNKGAAGGTTATGTACAGGAACGCAC-3′

GFPuvP89-R

5′-GTGCGTTCCTGTACATAACCTTCMNNCATGGCACTCTTGAAAAAGTCATG-3′

MjProRSC250G-F

5′- GATTATGCTTATCAAACAGGCTACGGAATCTCAGATAGGG-3′

MjProRSC250G-R

5′-CCCTATCTGAGATTCCGTAGCCTGTTTGATAAGCATAATC-3′

MjaG252T-F

5′- GATTATGCTTATCAAACAGGCTACACTATCTCAGATAGGGTTATAGC-3′

MjaG252T-R

5′-GCTATAACCCTATCTGAGATAGTGTAGCCTGTTTGATAAGCATAATC-3′

MjaG252A-F

5′- GATTATGCTTATCAAACAGGCTACGCTATCTCAGATAGGGTTATAGC-3′

MjaG252A-R

5′-GCTATAACCCTATCTGAGATAGCGTAGCCTGTTTGATAAGCATAATC-3′

MjaG252S-F

5′- GATTATGCTTATCAAACAGGCTACAGCATCTCAGATAGGGTTATAGC-3′

MjaG252S-R

5′-GCTATAACCCTATCTGAGATGCTGTAGCCTGTTTGATAAGCATAATC-3′

MjaG252C-F

5′- GATTATGCTTATCAAACAGGCTACTGCATCTCAGATAGGGTTATAGC-3′

MjaG252C-R

5′-GCTATAACCCTATCTGAGATGCAGTAGCCTGTTTGATAAGCATAATC-3′

MjProRSRM1-F

5′-GCATGGGTACCATGTTGGAATTTTCAGAATGG-3′

MjProRSRM1-R

5′-GCGATTCTAGATTAGTAGGTTTTAGCTATTGC-3′

SSMLMjProRS1-F (Mutation of E103)

5′-GCTTTAAGACCTACTTCANNKACACCAATATACTATATG-3′

SSMLMjProRS2-F (Mutation of F150 and E152)

5′-GGTTAAGAGAGATAATGACANNKAAANNKGCCCACACTGCCCATTC-3′

SSMLMjProRS3-F (Mutation of C249)

5′-GATTATGCTTATCAAACANNKTACGGAATCTCAGATAGGG-3′

Fluorescence-activated cell sorting (FACS) and analysis

FACS analysis and sorting of GFPuv and superfolder GFP containing cells were carried out on a FACSVantage, SE model, system manufactured by Becton Dickinson (BD Biosciences, San Jose, CA), using the UV (365 nm) split laser for excitation of GFPuv with a 510/20 filter and the 488 nm argon ion blue laser for excitation of superfolder GFP with the 530/40 bandpass filter. FACS analysis, of cells expressing superfolder GFP, was carried out using the LSR II system also by Becton Dickinson (BD Biosciences, San Jose, CA). The 488 nm laser was used for excitation of superfolder GFP with 505 nm and 530/30 nm emission filters.

Negative Sorting

UQ27$^{ts,\ ProA}$ cells expressing the fluorescent reporter (GFPuv or superfolder GFP) and pooled libraries of 6.0 x 10^5 variants of the MjProRS gene were grown in 5 mL LB supplemented with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL) overnight at 30 ºC with shaking at 225 rpm. A positive control (cells expressing the elastin-reporter fusion and pHC15 (MjPair), or cells expressing the fluorescent GFP alone) and a negative control (cells expressing the elastin-reporter fusion with no MjPair) were prepared with the experimental samples (cells expressing the elastin-reporter fusion and the MjProRS mutant library plasmids). Two mL of the overnight cultures were transferred to 48 mL SMM + proline liquid media (supplemented with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL) that had been added to 250 mL Erlenmeyer flasks for a total volume of 50 mL growth culture. The 50 mL cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a

period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), the 50 mL cultures were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles to spin down the cells.

The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were finally resuspended in 9 mL SMM - proline minimal media for each of 2 small duplicate flasks (18 mL SMM - proline total) for each cell type. The cell cultures were incubated at 42 °C for 30 min with shaking at 225 rpm to deplete the remaining cellular levels of proline.

The osmolarity of the culture was adjusted prior to induction via addition of the appropriate osmolytes to a final concentration of either 600 mM for NaCl. For example, 1 mL of a 5 M NaCl stock solution (29.22 g NaCl in 100 mL H_2O, sterilized) was added to each expression flask. In addition, proline analogues, such as 2*S*,4*R*)-4-hydroxyproline or (2*S*,4*R*)-4-fluoroproline, were prepared as 1 mL 200 mM stock solutions in PBS at pH 7.4 and stored at 4 °C. L-proline was prepared as a 1 mL 200 mM stock solution in PBS at pH 7.4 and stored at 4 °C. A 100 μL aliquot of the L-proline or the proline analogue stock solutions was added to the 10 mL cell cultures for a final concentration of 2 mM

proline or proline analogue. For negative screens, proline only was added to the cell culture before induction. To induce protein expression, 10 μL 1M IPTG were added to each flask for a final concentration of 1 mM IPTG. The IPTG 1000X (1M) stock was prepared by dissolving 596 mg IPTG in 2.5 mL ddH_2O followed by filter sterilization.

The cultures were incubated at 37 °C with shaking at 225 rpm for 3 h after induction. OD_{600} readings were monitored for each of the cultures. One mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 μL ddH_2O and stored at –20 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 μL sample culture, 12.5 μL 2 x sample buffer, 1 μL DTT (dithiothreitol), and 6.5 μL ddH_2O for a total of 25 μL. Samples were placed at 95 ºC for 5 min.

After expression, 1 mL aliquots of each cell type were collected in 1.5 mL Eppendorf tubes. The cell suspensions were stored overnight at 4 °C, and FACS analysis and sorting were conducted the following day. For FACS analysis and sorting, the cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C and the supernatant was discarded. The cell pellets were washed and redissolved in 10 mM phosphate-buffered saline (PBS) at pH 7.4. The cells were diluted with PBS to an OD_{600} of ~0.05. The cells were kept on ice and FACS sorting and analysis were carried out at ~4 °C. In all cases 2-4 rounds of negative sorting were carried out after the first expression. For negative sorting, the positive and negative controls as well as the experimental samples were analyzed first before sorting. For sorting itself, the throughput rate of cells was set at 5-10,000 events per second. Gates were set in the forward scatter and side scatter channels to exclude events arising from large particles. The gate in the fluorescence channel was

set to recover 0.3-0.5 % of the least fluorescent cells. Approximately 10^5 cells were sorted first in purity mode to ensure full coverage of the library. If sorting was especially slow, the enrich mode was used to collect the first population of cells. The sorted cells were then immediately subjected to sorting again under the same conditions (in purity mode). 10,000-500,000 events were collected from the enriched population. Approximately 4 mL SOC were added to the PBS in the tube containing the sorted cells and the cells were recovered for 1-2 h after sorting. The cells were then grown overnight after addition of antibiotics to the media or the suspension was plated for collection of the cells. Frozen stocks were prepared and stored at -80 ºC after each sort. The negative screening step was repeated 1-2 more times and analysis was used after sorting to determine the level of fluorescence of the new cell populations.

Positive Sorting

UQ27$^{ts,\ ProA}$ cells expressing the fluorescent reporter (GFPuv or superfolder GFP) and pooled libraries from previous negative sorting rounds (50 µL from a frozen stock) were grown in 5 mL LB supplemented with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL) overnight at 30 ºC with shaking at 225 rpm. A positive control (cells expressing the elastin-reporter fusion and pHC15 (MjPair), or cells expressing the fluorescent GFP alone) and a negative control (cells expressing the elastin-reporter fusion with no MjPair) were prepared with the experimental samples (cells expressing the elastin-reporter fusion and the MjProRS mutant library plasmids). Two mL of the overnight cultures were transferred to 48 mL SMM + proline liquid media (supplemented with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL) that had been added to 250 mL Erlenmeyer flasks for a total volume of 50

mL growth culture. The 50 mL cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), the 50 mL cultures were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles to spin down the cells.

The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were finally resuspended in 9 mL SMM - proline minimal media for each of 2 small duplicate flasks (18 mL SMM - proline total) for each cell type. The cell cultures were incubated at 42 °C for 30 min with shaking at 225 rpm to deplete the remaining cellular levels of proline.

The osmolarity of the culture was adjusted prior to induction via addition of the appropriate osmolytes to a final concentration of either 600 mM for NaCl. For example, 1 mL of a 5 M NaCl stock solution (29.22 g NaCl in 100 mL H_2O, sterilized) was added to each expression flask. In addition, proline analogues, such as (2*S*,4*R*)-4-hydroxyproline, were prepared as 1 mL 200 mM stock solutions in PBS at pH 7.4 and stored at 4 °C. L-proline was prepared as a 1 mL 200 mM stock solution in PBS at pH 7.4 and stored at 4 °C. A 100 µL aliquot of the L-proline or the proline analogue stock

solutions was added to the 10 mL cell cultures for a final concentration of 2 mM proline or 2 mM proline + 2 mM *trans*-4-hydroxyproline. For positive screens, proline only was added to one cell culture while a duplicate culture was induced in the presence of proline and hydroxyproline combined in one expression flask. To induce protein expression, 10 µL 1M IPTG were added to each flask for a final concentration of 1 mM IPTG. The IPTG 1000X (1M) stock was prepared by dissolving 596 mg IPTG in 2.5 mL ddH_2O followed by filter sterilization.

The cultures were incubated at 37 °C with shaking at 225 rpm for 3 h after induction. OD_{600} readings were monitored for each of the cultures. One mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 µL ddH_2O and stored at –20 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 µL sample culture, 12.5 µL 2 x sample buffer, 1 µL DTT (dithiothreitol), and 6.5 µL ddH_2O for a total of 25 µL. Samples were placed at 95 ºC for 5 min.

After expression, 1 mL aliquots of each cell type were collected in 1.5 mL Eppendorf tubes. The cell suspensions were stored overnight at 4 °C, and FACS analysis and sorting were conducted the following day. For FACS analysis and sorting, the cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C and the supernatant was discarded. The cell pellets were washed and redissolved in 10 mM phosphate-buffered saline (PBS) at pH 7.4. The cells were diluted with PBS to an OD_{600} of ~0.05. The cells were kept on ice and FACS sorting and analysis were carried out at ~4 °C. In all cases 2-4 rounds of positive sorting were carried out after negative sorting. For positive sorting, the positive and negative controls as well as the experimental samples were analyzed first

before sorting. For sorting itself, the throughput rate of cells was set at 5-10,000 events per second. Gates were set in the forward scatter and side scatter channels to exclude events arising from large particles. The gate in the fluorescence channel was set to recover 0.3-0.5 % of the least fluorescent cells. Approximately 10^5 cells were sorted first in purity mode to ensure full coverage of the library. If sorting was especially slow, the enrich mode was used to collect the first population of cells. The sorted cells were then immediately subjected to sorting again under the same conditions (in purity mode). 10,000-500,000 events were collected from the enriched population. Approximately 4 mL SOC were added to the PBS in the tube containing the sorted cells and the cells were recovered for 1-2 h after sorting. The cells were then grown overnight after addition of antibiotics to the media or the suspension was plated for collection of the cells. Frozen stocks were prepared and stored at -80 ºC after each sort. The positive screening step was repeated 1-2 more times and analysis was used after sorting to determine the level of fluorescence of the new cell populations.

Protein Expression

Small-scale expression in the UQ27$^{ts,\ ProA}$ strain*

Plasmid DNA (pQE-80L containing the **elastin-CCC** gene or the **elastin-CCC** fusion genes) was co-transformed with pWK1 plasmid DNA (containing a recombinant, wild-type *E. coli* prolyl-tRNA synthetase gene) or the empty plasmid, pHEC1 (lacking the synthetase/tRNA pair), into the UQ27$^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). Elastin or fusion plasmid DNA also was also co-transformed with pHC15 plasmid DNA (containing the recombinant, wild-type *M. jannaschii* prolyl-tRNA

synthetase/tRNA^{Pro} gene), or other appropriate plasmids, into the UQ27$^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). Five mL LB media, supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL), were inoculated with single colonies of UQ27$^{ts,\ ProA}$ cells, harboring the appropriate plasmids, and grown overnight at 30 °C on a rotator. Two mL of the overnight cultures were transferred to 48 mL LB (supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL) that had been added to 250 mL Erlenmeyer flasks for a total volume of 50 mL culture.

The 50 mL cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), protein expression was induced with IPTG. For induction, 50 uL, 1M IPTG were added to the flasks for a final concentration of 1 mM IPTG. The cultures were incubated at 42 °C for 3 h. OD_{600} readings were monitored for each of the cultures. 1 mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 μL ddH_2O and stored at –20 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 μL sample culture, 12.5 μL 2 x sample buffer, 1 μL DTT (dithiothreitol), and 6.5 μL ddH_2O for a total of 25 μL. Samples were placed at 95 ºC for 5 min.

*For expression in cell lines other than the UQ27 strain, the only modification from the outlined procedure is that cells were grown and incubated before and after induction at 37 ºC rather than at 30 ºC and then 42 °C.

Small-scale expression in minimal media for proline analogue incorporation

Plasmid DNA (pQE-80L containing the **elastin-CCC** gene or the **elastin-CCC** fusion genes) was co-transformed with pWK1 plasmid DNA (containing a recombinant, wild-type *E. coli* prolyl-tRNA synthetase gene) or the empty plasmid, pHEC1 (lacking the synthetase/tRNA pair), into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). Elastin or elastin fusion plasmid DNA was also co-transformed with pHC15 plasmid DNA (containing the recombinant, wild-type *M. jannaschii* prolyl-tRNA synthetase/$tRNA^{Pro}$ gene), or with other appropriate plasmids, into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). Five mL LB media, supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL), were inoculated with single colonies of $UQ27^{ts,\ ProA}$ cells, harboring the appropriate plasmids, and grown overnight at 30 °C on a rotator. A 2 mL aliquot of the overnight cultures was transferred to 48 mL NMM + proline or SMM + proline liquid media (supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL) that had been added to 250 mL Erlenmeyer flasks for a total volume of 50 mL growth culture.

The 50 mL cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), the 50 mL cultures were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles to spin down the cells. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The

cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 50 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were finally resuspended in 9 mL of NMM - proline or SMM - proline minimal media for each of 2 small duplicate flasks (18 mL NMM or SMM - proline total) for each cell type. The cell cultures were incubated at 42 °C for 30 min with shaking at 225 rpm to deplete the remaining cellular levels of proline.

For expressions involving proline analogues, the osmolarity of the culture was adjusted prior to induction via addition of the appropriate osmolytes to a final concentration of either 600 mM for NaCl or 800 mM for sucrose. For example, 1 mL of a 5 M NaCl stock solution (29.22 g NaCl in 100 mL H_2O, sterilized) was added to each expression flask. The osmolarity of the culture was adjusted for incorporation of all analogues with the exceptions of (2*S*,4*R*)-4-fluoroproline and (2*S*,4*S*)-4-fluoroproline which did not require osmotic shock for successful incorporation. In addition, proline analogues, such as (2*S*,4*R*)-4-hydroxyproline or (2*S*,4*R*)-4-fluoroproline, were prepared as 1 mL 200 mM stock solutions in PBS at pH 7.4 and stored at 4 °C. L-proline was prepared as a 1 mL 200 mM stock solution in PBS at pH 7.4 and stored at 4 °C. A 100 μL aliquot of the L-proline or the proline analogue stock solutions was added to the 10 mL cell cultures for a final concentration of 2 mM proline or proline analogue. For

positive screening using the **elastin-CCC**-GFPuv fusion reporters, proline was added to one aliquot of the cell culture (2 mM final concentration) while a duplicate flask contained cells exposed to proline and hydroxyproline combined at final concentrations of 2 mM each. For negative screens, proline only was added to the cell culture before induction. To induce protein expression, 10 µL 1M IPTG were added to each flask for a final concentration of 1 mM IPTG. The IPTG 1000x (1M) stock was prepared by dissolving 596 mg IPTG in 2.5 mL ddH_2O followed by filter sterilization.

The cultures were incubated at 42 °C with shaking at 225 rpm for 3 h after induction. OD_{600} readings were monitored for each of the cultures. One mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 µL ddH_2O and stored at –20 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 µL sample culture, 12.5 µL 2 x sample buffer, 1 µL DTT (dithiothreitol), and 6.5 µL ddH_2O for a total of 25 µL. Samples were placed at 95 ºC for 5 min.

Large-scale expression

Plasmid DNA containing the superfolder GFP expression plasmid was transformed into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 µg/mL) and ampicillin (100 µg/mL). Five milliliters (mL) LB media, supplemented with kanamycin (50 µg/mL) and ampicillin (100 µg/mL) were inoculated with single colonies of $UQ27^{ts,\ ProA}$ cells, harboring the appropriate plasmids, and grown overnight at 30 °C on a rotator. Five mL of the overnight cultures were transferred to 495 mL LB (supplemented with kanamycin (50 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL) that had been added to 2,800 mL Erlenmeyer flasks for a

total volume of 500 mL culture. Two flasks were used to express 1000 mL (1L) total volume of culture.

Once the cells reached log phase growth (OD_{600} ~0.7), the 500 mL cultures were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles to spin down the cells. The supernatant was discarded, and the cell pellets were resuspended in 500 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 500 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were resuspended in 500 mL aqueous, sterile cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4000 x g for 10 min, at 4 °C, in medium centrifuge bottles. The supernatant was discarded, and the cell pellets were finally resuspended in 500 mL SMM - proline minimal media. The cell cultures were incubated at 42 °C for 30 min with shaking at 225 rpm to deplete the remaining cellular levels of proline.

For expressions involving proline analogues, the osmolarity of the culture was adjusted prior to induction via addition of the appropriate osmolytes to a final concentration of either 600 mM for NaCl or 800 mM for sucrose. For example, 50 mL of a 5 M NaCl stock solution (29.22 g NaCl in 100 mL H_2O, sterilized) was added to each expression flask. The osmolarity of the culture was adjusted for incorporation of all analogues with the exceptions of (2*S*,4*R*)-4-fluoroproline and (2*S*,4*S*)-4-fluoroproline which did not require osmotic shock for successful incorporation. In addition, proline analogues, such as (2*S*,4*R*)-4-hydroxyproline or (2*S*,4*R*)-4-fluoroproline, were prepared

as 1 mL 200 mM stock solutions in PBS at pH 7.4 and stored at 4 °C. L-proline was prepared as a 1 mL 200 mM stock solution in PBS at pH 7.4 and stored at 4 °C. L-proline or the proline analogue stock solutions were added to the 10 mL cell cultures for a final concentration of 2 mM proline or proline analogue. For induction, 5 mL, 1M IPTG were added to the flasks for a final concentration of 1 mM IPTG. The cultures were incubated at 42 °C for 4 h. OD_{600} readings were monitored for each of the cultures. One mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 uL ddH_2O and stored at –20 °C. After 3 h of induction, the two large cultures (1 L total volume) were spun down at 4,000 x g for 20 min. The cell pellets were resuspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and stored at –80 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 µL sample culture, 12.5 µL 2 x sample buffer, 1 µL DTT (dithiothreitol), and 6.5 µL ddH_2O for a total of 25 µL. Samples were placed at 95 ºC for 5 min.

Protein purification (GFPuv or superfolder GFP)

Large-scale expressions of GFPuv or superfolder GFP were carried out on a 1 L scale. The expressions were conducted according to the procedure described above with incorporation of L-proline only or global incorporation of (2*S*,4*R*)-4-hydroxyproline only for each fluorescent reporter protein. Purification of GFP was facilitated by an N-terminal decahistidine tag in the pQE-80L vector. A large-scale expression was carried out to obtain a large amount of the GFP protein in the expression strain, $UQ27^{ts,\ ProA}$. The frozen cells were lysed by three freeze/thaw cycles. Lysozyme (1 mg/mL), EDTA-free protease inhibitor cocktail, benzonase (25 units/mL), and $MgCl_2$ (1 mM) were added to

the lysate and the mixture was incubated with shaking (225 rpm) at 4 °C, overnight. The cell lysate was centrifuged at 10,000 x *g* for 30 min at 4 °C. The supernatant and pellet were separated and analyzed by SDS-PAGE to determine the location of the target protein. SDS-PAGE analysis indicated that the majority of the target protein was present in the soluble fraction with some residual protein in the pellet. For the soluble elastin-mimetic protein, the supernatant was loaded onto TALON® Co^{2+} metal affinity resin (5 mL) and washed with lysis buffer (50 mL) containing 20 mM imidazole. The target protein was eluted with elution buffer (20 mL, 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.0) and dialyzed (MWCO = 10 kD) against distilled deionized water (5 × 4 L). The dialysate was lyophilized to produce a green/yellow solid.

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS)

The molar masses of superfolder GFP and variant superfolder GFP were determined by MALDI-TOF MS on an Applied Biosystems Voyager System 428 mass spectrometer in the positive linear mode. The matrix, 2-(4-hydroxyphenylazo)benzoic acid (HABA), was used at a concentration of 10 mg/ml in a 50:50 mixture of water and 2-propanol. The protein solution (1 mg/ml in ddH_2O) was mixed with the matrix solution in a ratio of 1:10 and dried under vacuum or air. Bovine serum albumin was used as a standard for external calibration. Sample preparation was repeated in triplicate and the average mass was calculated for each of 3 trials of MALDI-TOF-MS mass analysis. FT-MS with electrospray ionization was also used to determine the m/z ratios of the purified superfolder GFP protein with proline and hydroxyproline incorporation.

Results and Discussion

Site-directed mutagenesis of MjProRS

Several site-mutants of MjProRS were generated using site-directed mutagenesis to investigate substrate specificity of the enzyme. An EcProRS mutant enzyme was previously shown to accommodate a bulky proline analogue after only a single amino acid change in the amino acid sequence of the enzyme. The C443G mutant enzyme enabled global incorporation of (2*S*)-piperidine-2-carboxylic acid into elastin-mimetic proteins, while the wild-type enzyme was unable to accommodate the bulky substrate. A similar approach was taken for the MjProRS enzyme. The residues, C249 and G251, were chosen as targets for rational design of an enzyme with a preference for hydroxyproline over the natural substrate, proline. The C249 position is homologous to the C443 residue in the EcProRS enzyme that has been implicated as a structurally critical residue in the active site of the EcProRS[12]. Mutation of the C443 residue to a glycine residue enabled activity of the mutant EcProRS enzyme with a sterically demanding proline analogue that was not accommodated by the wild-type enzyme.

The positions S288 of the TtProRS and C265 of the MtProRS (*M. thermautotrophicus* ProRS) are also homologous to the C249 position in the MjProRS enzyme and are observed to make relatively close contacts between the side chain atoms and the *exo* face of the proline ring. The G251 position was also chosen as a target for rational design as glycine, with a sterically permissive hydrogen side chain, also made close contact (within 3 Å) with atoms in the proline ring (C4 atom) and could potentially be chemically altered to accommodate or induce more favorable binding and activity of the enzyme with hydroxyproline or with other proline analogues. For example, mutation

of C249 to a glycine was suspected to allow for accommodation of a bulky proline analogue, (2*S*)-piperidine-2-carboxylic acid in the active site of the enzyme as was the case for EcProRS in a previous study[13].

The wild type MjProRS as well as the C249G mutant enzyme enabled incorporation of (2*S*)-piperidine-2-carboxylic acid in **elastin-CCC** according to western blot analysis. A series of MjProRS mutants were generated to further evaluate substrate preference in the enzyme. The MjProRS mutant, C249G, showed activity for L-proline and for (2*S*,4*R*)-4-fluoroproline incorporation into **elastin-CCC**. The MjProRS double mutant, C249G, G251A enabled incorporation of L-proline but not with proline analogues. The other MjProRS double mutants (MjProRS C249G, G251T, double mutant MjProRS C249G, G251S, and double mutant MjProRS C249G, G251C) did not support expression of the elastin derivative with proline or with any of the proline analogues. Rational design did not result in any change in proline analogue substrate specificity compared to that of the wild type enzyme. A high-throughput screening system was then developed for engineering substrate specificity in the MjProRS enzyme.

Elastin-CCC/GFPuv fusion reporter systems

Using primers designed for site-directed mutagenesis, the CCC proline codon sites (P89 and P211) in the GFPuv gene (in the pGFPuv plasmid) were both converted to CCG sites producing silent mutations where each site would still encode proline but the presence of the CCC codons would not interfere with expression of the reporter. Two constructs were investigated to determine whether an N or C-terminal fusion of GFPuv to **elastin-CCC** would work best for screening purposes. Both fusion constructs resulted in expression of full-length **elastin-CCC**-GFPuv and GFPuv-**elastin-CCC** proteins.

However, the **elastin-CCC**-GFPuv construct was chosen for screening as the background expression levels with the C-terminal fusion were lower that those with the N-terminal GFPuv fusion. The **elastin-CCC**-GFPuv was also desirable because it could be used to demonstrate read-through of the **elastin-CCC** gene, containing ~80 CCC sites, and subsequent production of the reporter protein fused to the elastin gene. The level of fluorescence was low compared to the GFPuv-**elastin-CCC** protein, possibly due to insolubility and/or aggregation of the **elastin-CCC** protein portion that interfered with GFPuv folding/fluorescence. FACS analysis was conducted to determine baseline levels of fluorescence of cells expressing the **elastin-CCC**-GFPuv fusion reporter protein with co-expression of the MjProRS/MjtRNAPro pair as compared to no pair (empty plasmid). A higher level of fluorescence was detected when the orthogonal pair was present, but fluorescence levels were low compared to expression of the GFPuv protein alone (Fig. 5).

In Figure 5 (**A.**)**,** the negative control (red) indicated very low to no fluorescence when proline was not added to the media with expression of the reporter in the proline auxotrophic strain, UQ27$^{ts,\ ProA}$. The background level expression (brown) was almost an order of magnitude lower in fluorescence compared to the fluorescence of cells expressed with the orthogonal pair (blue). Expression of the fusion reporter resulted in a lower level of fluorescence overall compared to GFPuv expression (green). Although the overall level of fluorescence was slightly increased with the N-terminal fusion of GFPuv to **elastin-CCC** (pHC44), the level of fluorescence was still lower than that of GFPuv alone (green) as seen in Figure 5 (**B.**). Also, there was no obvious difference between the level of fluorescence of the background level expression (brown) and the level of fluorescence with the pair present (blue).

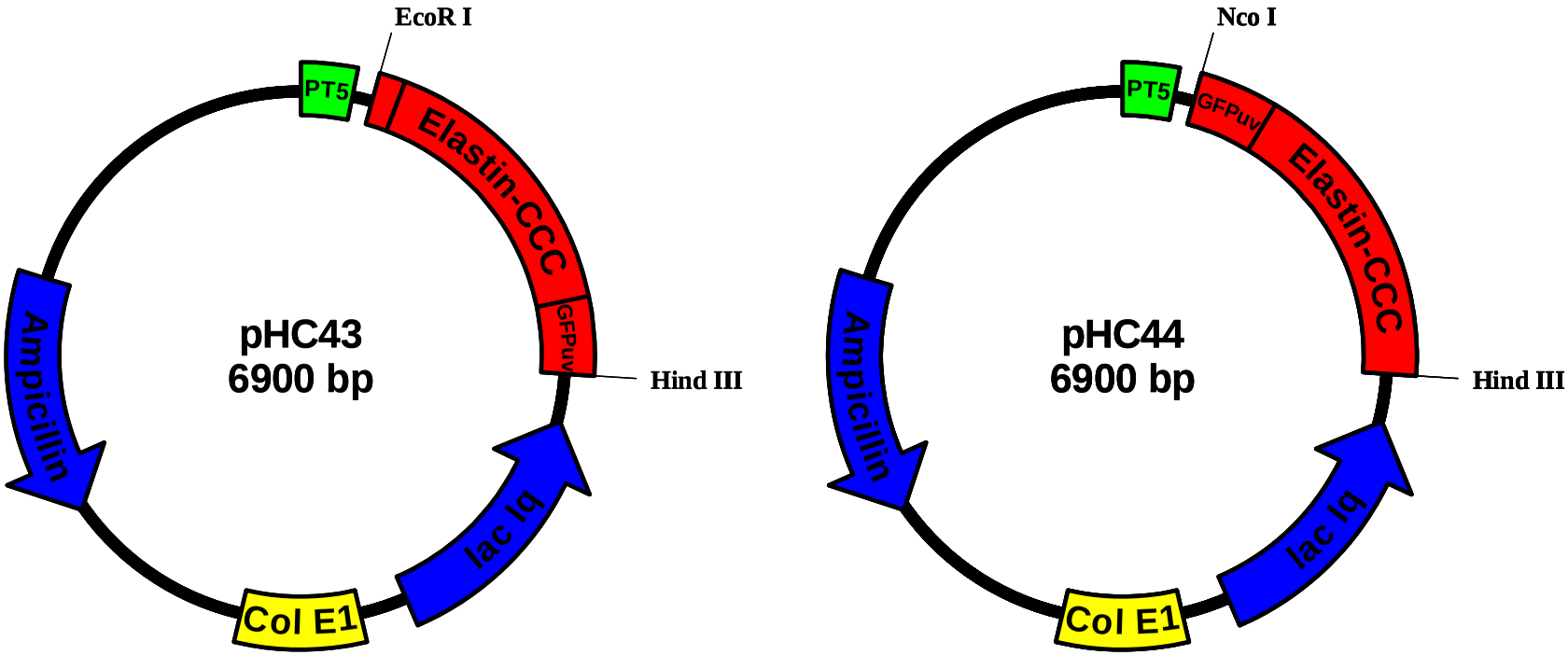


Figure 4. Plasmids for the expression of fusions of **elastin-CCC** to GFPuv. The pHC43 plasmid was generated by cloning the GFPuv gene into the pQE-80L expression plasmid harboring the **elastin-CCC** gene cassette to generate the C-terminal **elastin-CCC**-GFPuv fusion reporter. The pHC44 plasmid was generated by cloning the GFPuv gene into the pQE-80L expression plasmid harboring the **elastin-CCC** gene cassette to generate the N-terminal GFPuv-**elastin-CCC** fusion reporter.

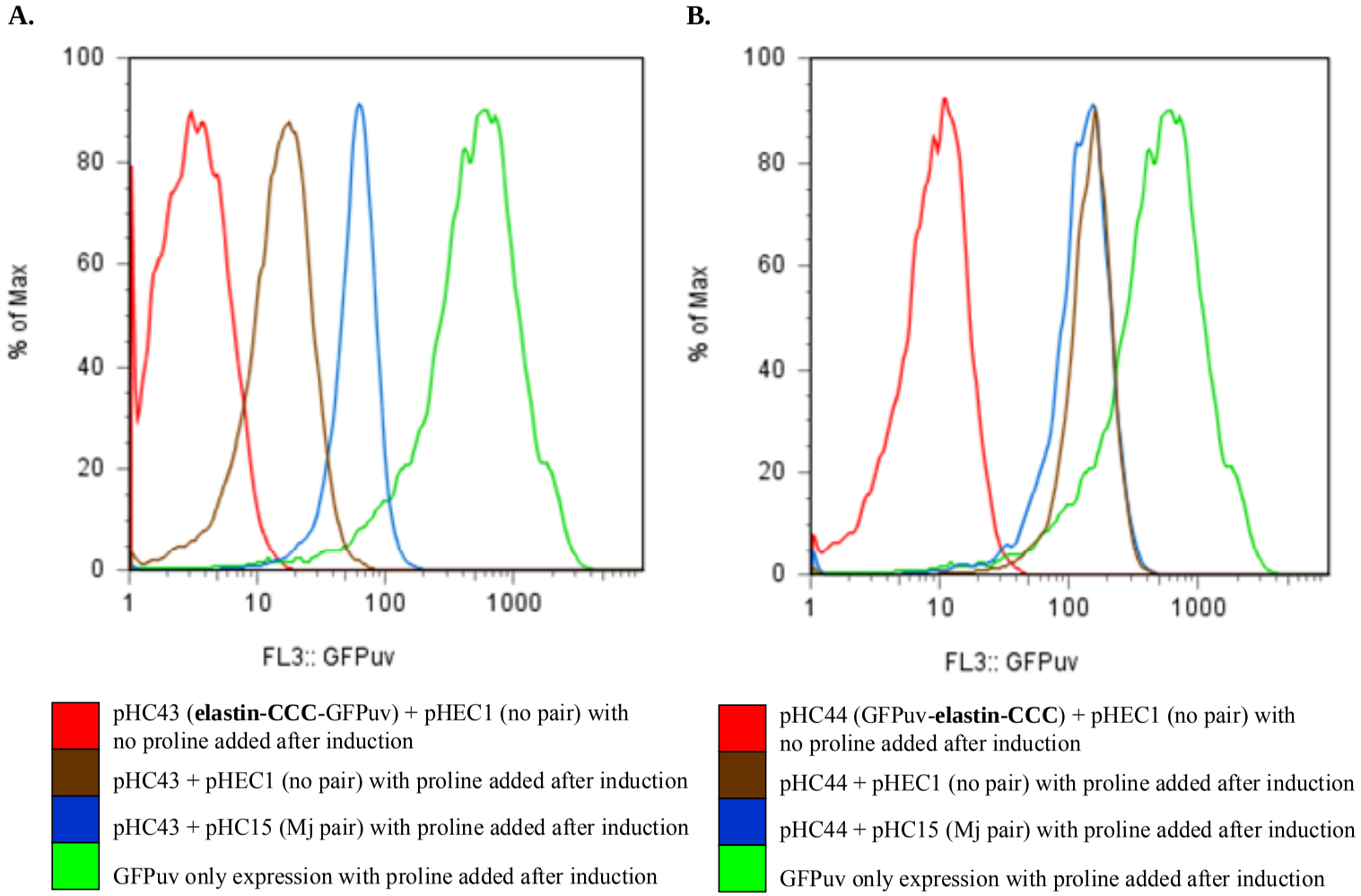


Figure 5. FACS analysis of cells after expression of either pHC43 (**elastin-CCC**-GFPuv) or pHC44 (GFPuv-**elastin-CCC)** in the presence of proline at 37 ºC. Expression of the fluorescent reporter proteins was carried out with co-expression of either the plasmid containing no orthogonal pair or the plasmid containing the MjProRS/MjtNRA^{Pro} pair.

Random mutagenesis and screening of MjProRS

Random mutagenesis of MjProRS was investigated as an initial method for mutagenesis strategies. Random mutagenesis of enzymes is often based on error-prone PCR amplification strategies, and random mutagenesis has been used in a variety of applications to alter the activity of enzymes. In some cases, the residues that are altered are not predicted to have a great impact on substrate specificity or activity of the enzyme and may lie outside of the active site. An essentially random technique allows for alteration of amino acid residues outside the active site that may play an important role in binding of the substrate or activity of the enzyme in the context of the 3D structure. Frances Arnold, in particular, has found success in targeting regions outside the active site for mutagenesis and directed evolution techniques used for altering enzyme activity. It is often necessary to carry out several rounds of random mutagenesis and screening to achieve altered substrate specificity in enzymes of interest.

A random library of MjProRS mutants was generated using Stratagene's GeneMorph II Random Mutagenesis Kit. The error-prone polymerase enabled mutagenesis of the synthetase at high and medium mutagenesis frequencies depending on how much template DNA was added to the reaction. The random library was screened by sequencing and the DNA was transformed into the expression strain containing the **elastin-CCC**-GFPuv fusion reporter. FACS analysis revealed a generally low level of fluorescence in the case of cells expressing the Mj pair as well as in cells expressing no pair. Cells expressing the reporter with the Mj pair exhibited slightly higher fluorescence levels as compared to the negative control. After three rounds of sorting of the random library, no difference was seen between cells expressing the random library and

fluorescent reporter in the presence of proline vs. those expressing in the presence of proline and hydroxyproline combined (Fig. 6). The random mutagenesis technique was not aggressively pursued as site-saturation mutagenesis was shown later to yield more promising results and was a highly successful method used in previous studies as a first approach for generation of synthetase enzymes with altered substrate specificities.

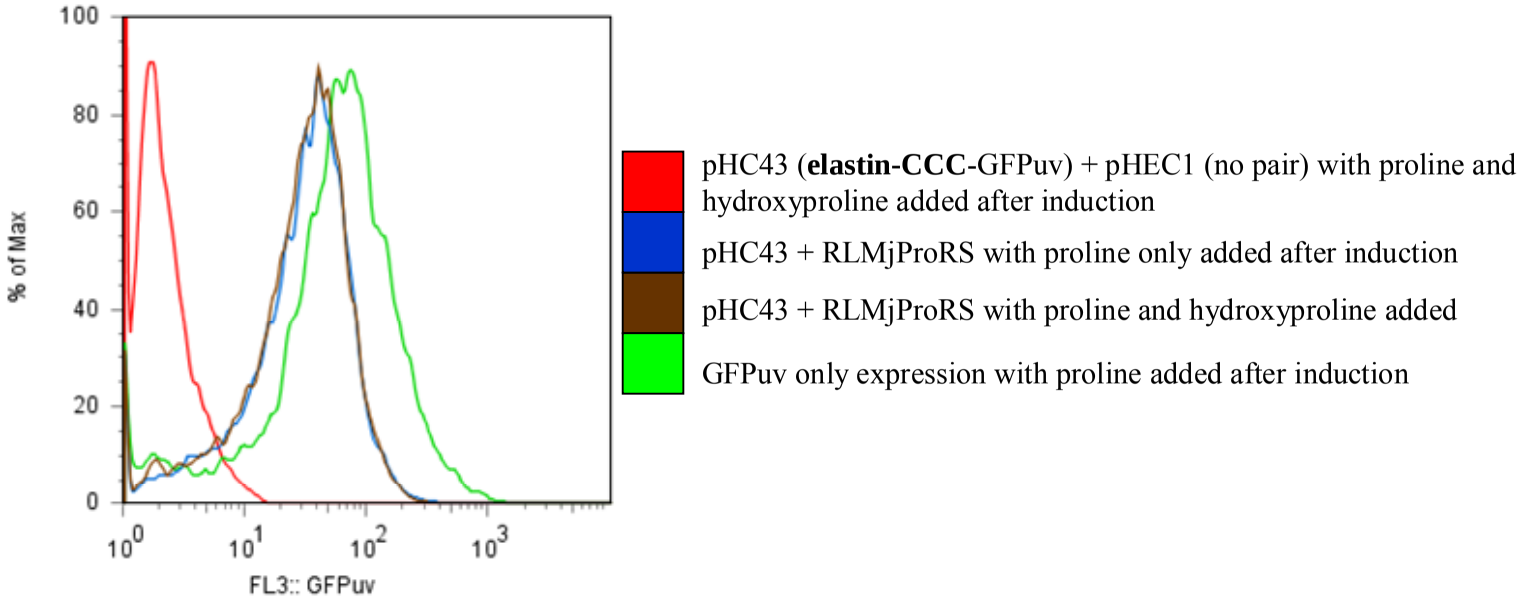


Figure 6. Results of sorting the random library of MjProRS variants after three rounds of sorting. FACS analysis of cells after expression of pHC43 (**elastin-CCC**-GFPuv) and the random library of MjProRS enzymes in the presence of proline or proline and hydroxyproline at 37 ºC. Expression of the fluorescent reporter proteins was carried out with co-expression of either the plasmid containing no orthogonal pair or the plasmid containing the mutant MjProRS/MjtNRA^{Pro} pairs. There was no difference between cells expressing the random library and reporter protein in the presence of proline (blue) vs. those expressing in the presence of proline and hydroxyproline (brown).

Site-saturation mutagenesis and screening of MjProRS

Site-saturation mutagenesis was pursued as a more targeted method for altering the substrate specificity of the MjProRS enzyme. Based on examination of the crystal structure of the homologous ProRS enzyme from *T. thermophilus* (PDB code: 1H4T), four residues within the active site of the MjProRS enzyme were chosen for site-saturation mutagenesis (E103, F150, E152, and C249). Each of the four residues chosen as targets for altering substrate specificity of the enzyme made close contacts (1.5-3.5 Å) with the bound L-proline substrate in the TtProRS homolog. Although the crystal structure of the MjProRS enzyme is known (PDB code: 1NJ8), the structure does not include a bound substrate. ClustalW analysis revealed high homology between the proteins at the primary sequence level (Appendix 1). Overlay of the MjProRS and TtProRS enzyme crystal structures using PyMOL software also revealed high homology between the two structures at the 2° and 3° levels. One limitation in using the PyMOL program was the inability to overlay the two structures while maintaining the presence of the bound substrate in the TtProRS structure.

The first attempt at site-saturation mutagenesis of MjProRS resulted in a low mutagenesis frequency at all four sites of interest. The library had a theoretical size of 160,000 (20^4) possible permutations given the 4 sites for saturation and 20 possible amino acids to occupy each site. The theoretical library size was just covered after several attempts at high-efficiency transformation into the expression strain already harboring the reporter protein plasmid as ~200,000 colonies were obtained and collected for screening with the **elastin-CCC**-GFPuv reporter. Obtaining high transformation efficiency was a particularly difficult part of the directed evolution procedure and

required a great deal of optimization. The strain of interest, containing an endogenous mutation and the two expression plasmids, was under a great deal of stress as growth of the strain required three different antibiotics and a modification of the growth and expression temperatures. Also, the frequency of mutagenesis was likely low due to the method of library generation at the initial cloning step. Initially, mutant synthetase enzymes were cloned into the expression vector by digestion of pHC15 (already containing the wild type MjProRS) and subsequent cloning of the mutants in place of the wild type gene. It is likely that uncut or incomplete digestion of the plasmids resulted in a higher number of wild type genes in the mutant pool. To avoid this on the second attempt at site-saturation mutagenesis, the MjProRS library (DNA from the mutagenesis reaction) was purified by butanol precipitation and directly transformed into high-efficiency Top10F´ electrocompetent cells immediately after mutagenesis (see methods).

After transformation, the library was initially subjected to 2 rounds of negative screening and 2 rounds of positive screening using FACS. Each round consisted of 2 actual sorting events where the first sort was conducted to enrich for clones of the desired fluorescence level and the second sort was used to further purify the most highly fluorescent cells or the cells with the lowest fluorescence in the population. 92 colonies were obtained after the final 2 rounds of positive sorting and they were individually grown and screened in 96-well plates in triplicate. Each colony was subjected to expression in the presence of proline or hydroxyproline and fluorescence levels were determined using a fluorescence plate reader. The results of the initial screen were highly ambiguous as the fluorescence results from a single colony in triplicate were not similar

in the majority of cases. Reproducibility of the results in the 96-well plate format using the plate reading spectrophotometer was poor.

Each of the 92 colonies was expressed individually according to the small-scale expression protocol in small flasks as the 96-well plate format was not ideal for the reporter system used. Ten of the most promising mutants were selected for individual expression to evaluate expression levels with proline or hydroxyproline only using the **elastin-CCC** protein only (not the fusion) as a reporter to confirm the results of the fluorescence-based evaluation. Surprisingly, each of the ten promising mutants selected from the fluorescence-based assay exhibited either a lower elastin-CCC expression level with proline and hydroxyproline compared to proline only (Fig. 9) Some mutants exhibited wild type expression profiles with expression levels of elastin-CCC approximately the same in the presence of proline or proline and hydroxyproline (Fig. 9). This was contrary to the expected result based on the screening strategy. Also, we were unable to sequence the MjProRS genes after many trials and several attempted sequencing strategies.

One possibility for the unexpected result may have been related to the initial method for cloning the site-saturation library. With several wild type enzymes present in the naïve library, this may explain why some of the clones isolated exhibited wild type expression levels. Also, FACS has been optimized as a technique for sorting eukaryotic cells and the screening system involved sorting of, much smaller, prokaryotic cells expressing a weakly fluorescent **elastin-CCC**-GFPuv fusion reporter. Heterogeneous fluorescence between cells was possible even with viable MjProRS mutants due to the fusion of the insoluble **elastin-CCC** to GFPuv. Improper folding of GFPuv in the fusion

protein could have contributed to low fluorescence levels thereby limiting the effectiveness of the FACS sorting technique.

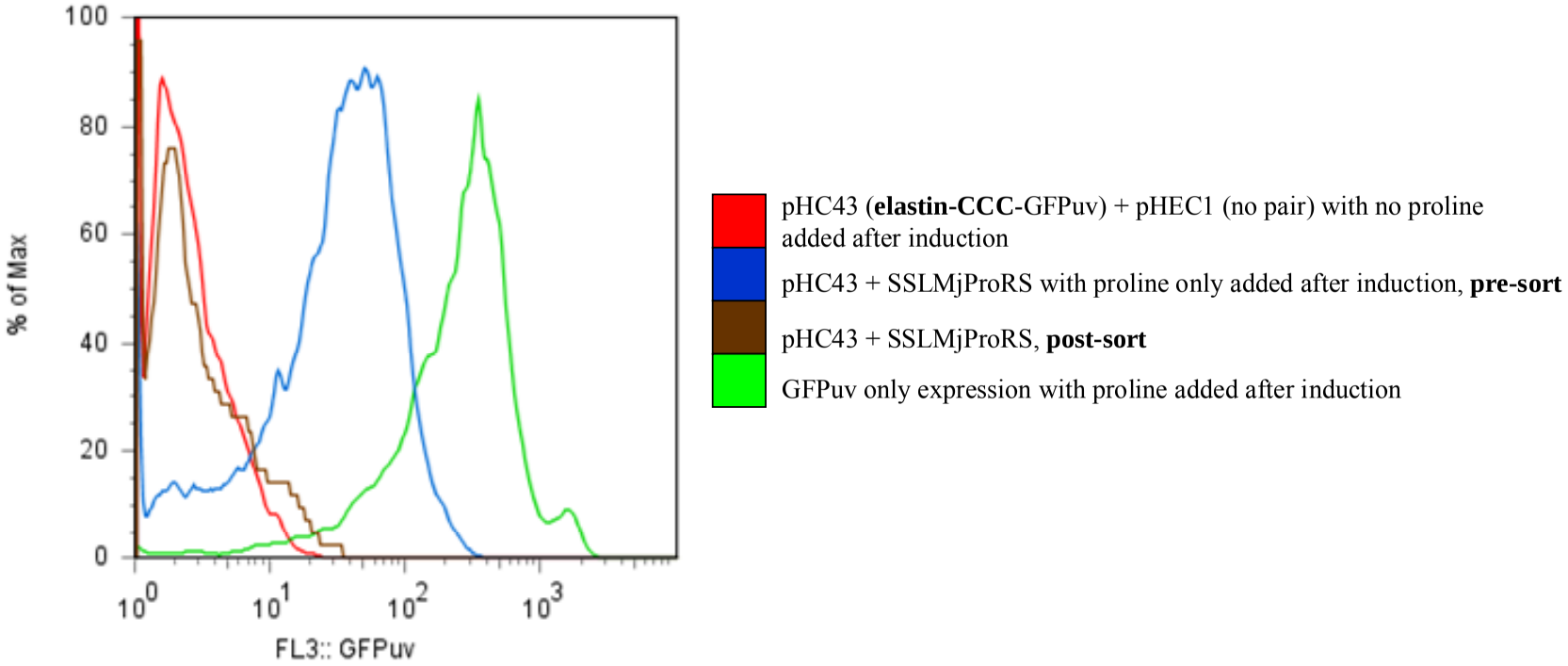


Figure 7. Example of one round of negative sorting of the MjProRS site-saturation mutant library with FACS using the fluorescent reporter, pHC43 (**elastin-CCC**-GFPuv). Expression of the fluorescent reporter protein was carried out with co-expression of the site-saturation library of MjProRS variants. The negative control (red) indicated very low to no fluorescence when proline was not added to the media with expression of the reporter in the proline auxotrophic strain, UQ27$^{ts, ProA}$. The starting level of cellular fluorescence (blue) was high compared to the fluorescence of cells detected after sorting for the most negative cells (brown). GFPuv expression alone is shown in green.

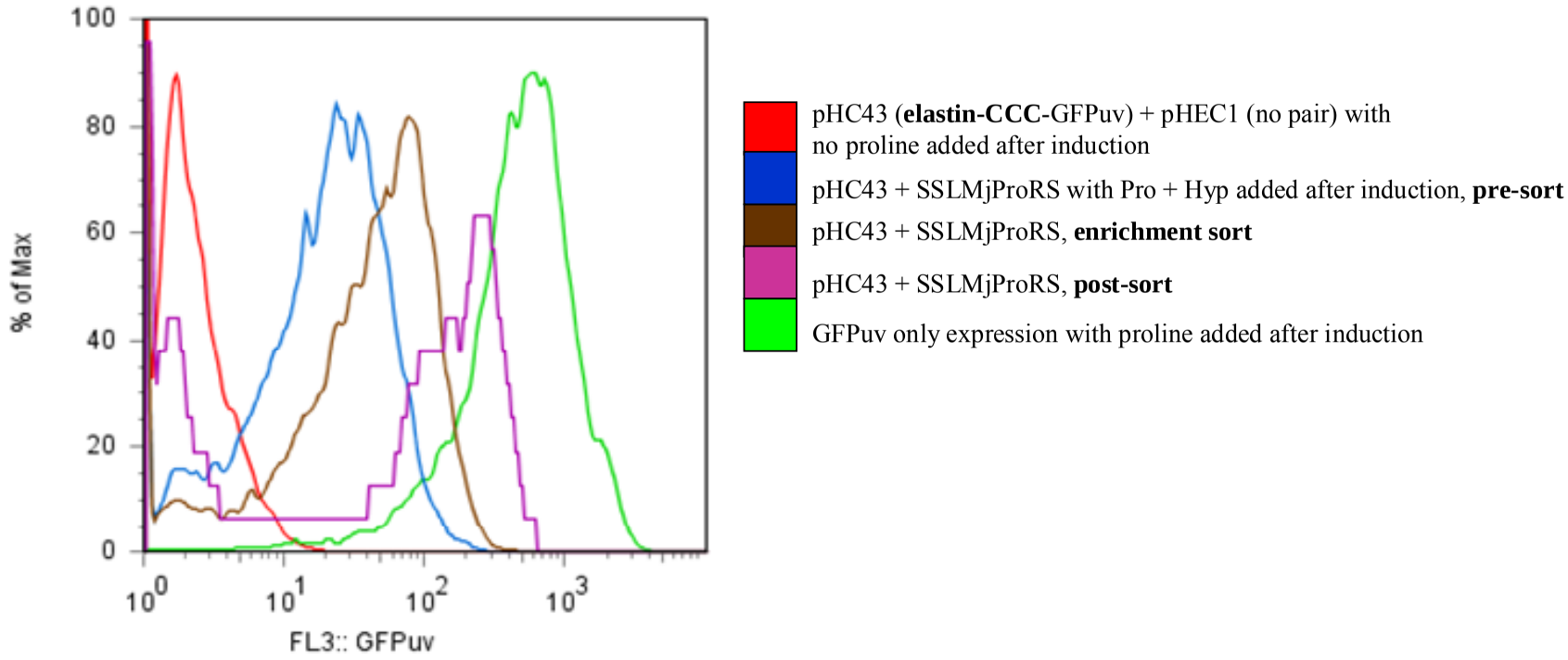
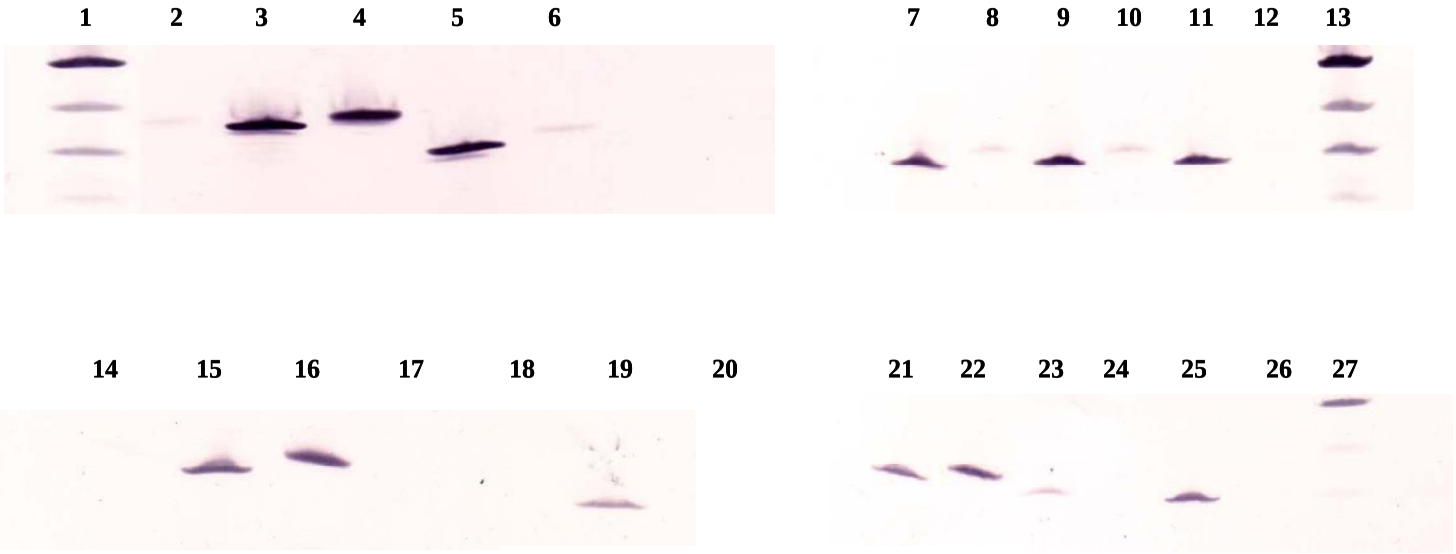


Figure 8. Example of one round of positive sorting of the MjProRS site-saturation mutant library with FACS using the fluorescent reporter, pHC43 (**elastin-CCC**-GFPuv). Expression of the fluorescent reporter protein was carried out with co-expression of the site-saturation library of MjProRS variants. The negative control (red) indicated very low to no fluorescence when proline was not added to the media with expression of the reporter in the proline auxotrophic strain, UQ27$^{ts,\ ProA}$. The starting level of cellular fluorescence (blue) was analyzed and a gate was drawn around the most highly fluorescent cells. The enrichment step (brown) indicated that the cells that were sorted were enriched for a higher level of fluorescence. The second sorting in purity mode allowed for further enrichment of highly fluorescent cells (purple). GFPuv expression alone is shown in green.



1. Perfect protein marker
2. pHC8 + pHEC1 (no pair), + hydroxyproline
3. pHC8 + pHC15 (Mj pair), + proline
4. pHC8 + pHC15 (Mj Pair), + hydroxyproline
5. pHC8 + Mutant 18, + proline
6. pHC8 + Mutant 18, + hydroxyproline
7. pHC8 + Mutant 20, + proline
8. pHC8 + Mutant 20, + hydroxyproline
9. pHC8 + Mutant 21, + proline
10. pHC8 + Mutant 21, + hydroxyproline
11. pHC8 + Mutant 22, + proline
12. pHC8 + Mutant 22, + hydroxyproline
13. Perfect protein marker
14. Perfect protein marker
15. pHC8 + Mutant 25, + proline
16. pHC8 + Mutant 25, + hydroxyproline
17. pHC8 + Mutant 39, + proline
18. pHC8 + Mutant 39, + hydroxyproline
19. pHC8 + Mutant 52, + proline
20. pHC8 + Mutant 52, + hydroxyproline
21. pHC8 + Mutant 56, + proline
22. pHC8 + Mutant 56, + hydroxyproline
23. pHC8 + Mutant 74, + proline
24. pHC8 + Mutant 74, + hydroxyproline
25. pHC8 + Mutant 90, + proline
26. pHC8 + Mutant 90, + hydroxyproline
27. Perfect protein marker

Figure 9. Western blot analysis indicating results of flask-based expressions of 10 mutant MjProRS enzymes isolated from FACS sorting. The plasmids were isolated by miniprep and transformed into the UQ27 expression strain already containing pHC8 (**elastin-CCC**). Mutant synthetases 18, 20, 21, 22, 52, 74, and 90 seem to favor activity with proline over hydroxyproline. Mutants 25 and 56 show activity with both proline and hydroxyproline, and mutant 39 appears to be inactive with both substrates. None of the mutants isolated showed a preference for hydroxyproline after several rounds of sorting.

A second site-saturation library was constructed and transformed into the expression strain. Successful mutagenesis was attributed to the following changes. The second attempt at site-saturation mutagenesis of MjProRS resulted in a high mutagenesis frequency at all four sites of interest as the library (DNA from the mutagenesis reaction) was purified by butanol precipitation and directly transformed into high-efficiency Top10F′ electrocompetent cells immediately after mutagenesis (Table 1). The DNA was isolated from the cells using plasmid miniprep to isolate a concentrated sample of the plasmid DNA harboring the mutant library genes. The library was then, subsequently, transformed into the UQ27 expression strain. The theoretical library size (1.6×10^5) was more than covered after high-efficiency transformation into the expression strain already harboring the reporter protein plasmid as ~600,000 colonies were obtained and collected for screening of the mutant library with the **elastin-CCC**-GFPuv reporter.

After transformation, the library was initially subjected to 2 rounds of negative sorting and 3 rounds of positive sorting using FACS. Each round consisted of 2 actual sorting events where the first sort was conducted to enrich for clones of the desired fluorescence level and the second sort was used to further purify the most highly fluorescent cells or the cells with the lowest fluorescence in the population. Within-day sorting indicated successful enrichment of either the negative or highly fluorescent cells as before. However, after several rounds of day-to-day sorting, only a slight difference was seen between the total populations of cells expressing the reporter with proline only or proline/hydroxyproline combined perhaps due to the overall low level of fluorescence of the **elastin-CCC**-GFPuv fusion reporter (Fig. 10). The insoluble **elastin-CCC** component of the fusion could have interfered with proper GFPuv folding resulting in a

decrease in the overall fluorescence level of the cells. Individual clones were not selected for further screening by western blot as the new reporter system, containing superfolder GFP, indicated significant promise in use as a more efficient reporter for FACS-based screening.

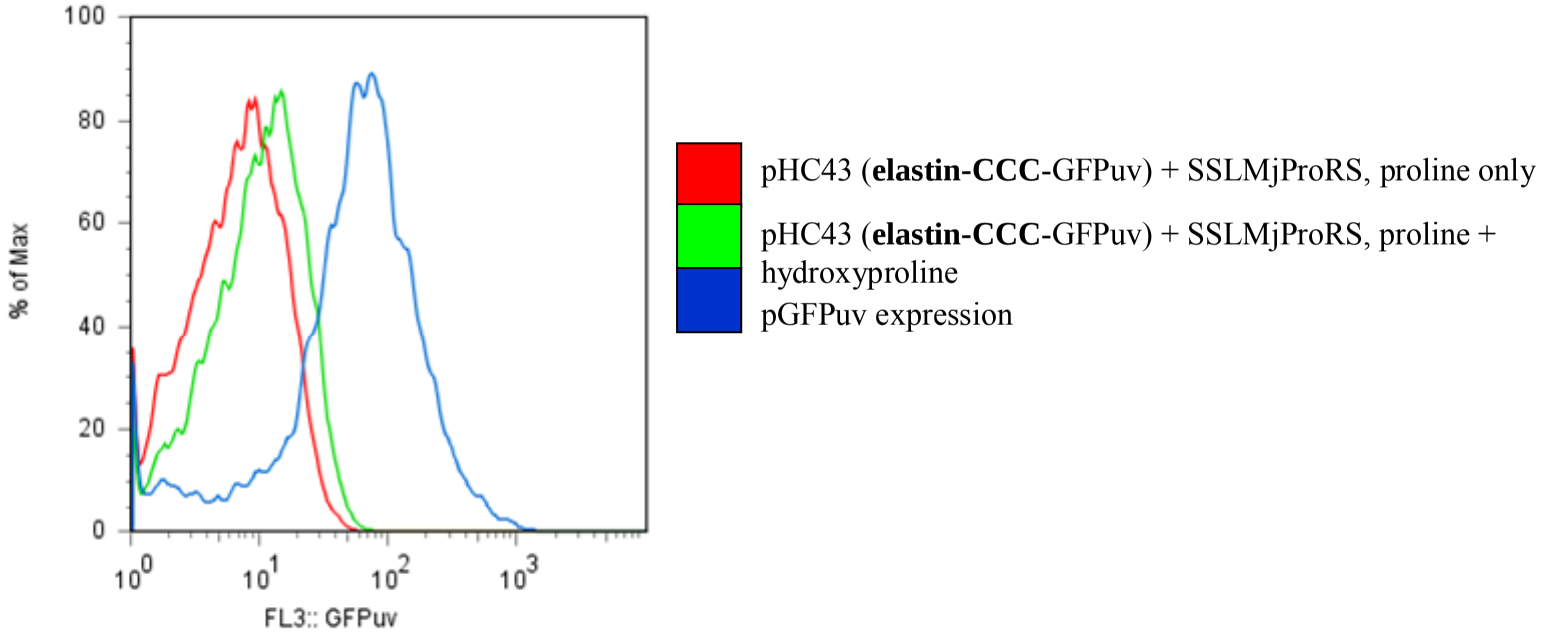


Figure 10. FACS analysis depicting results from screening of the site-saturation library of MjProRS mutants after 2 rounds of negative sorting followed by 3 rounds of positive sorting. Protein expression was induced at 37 °C in media containing either proline only (blue and red traces) or a mixture of proline and hydroxyproline (green trace).

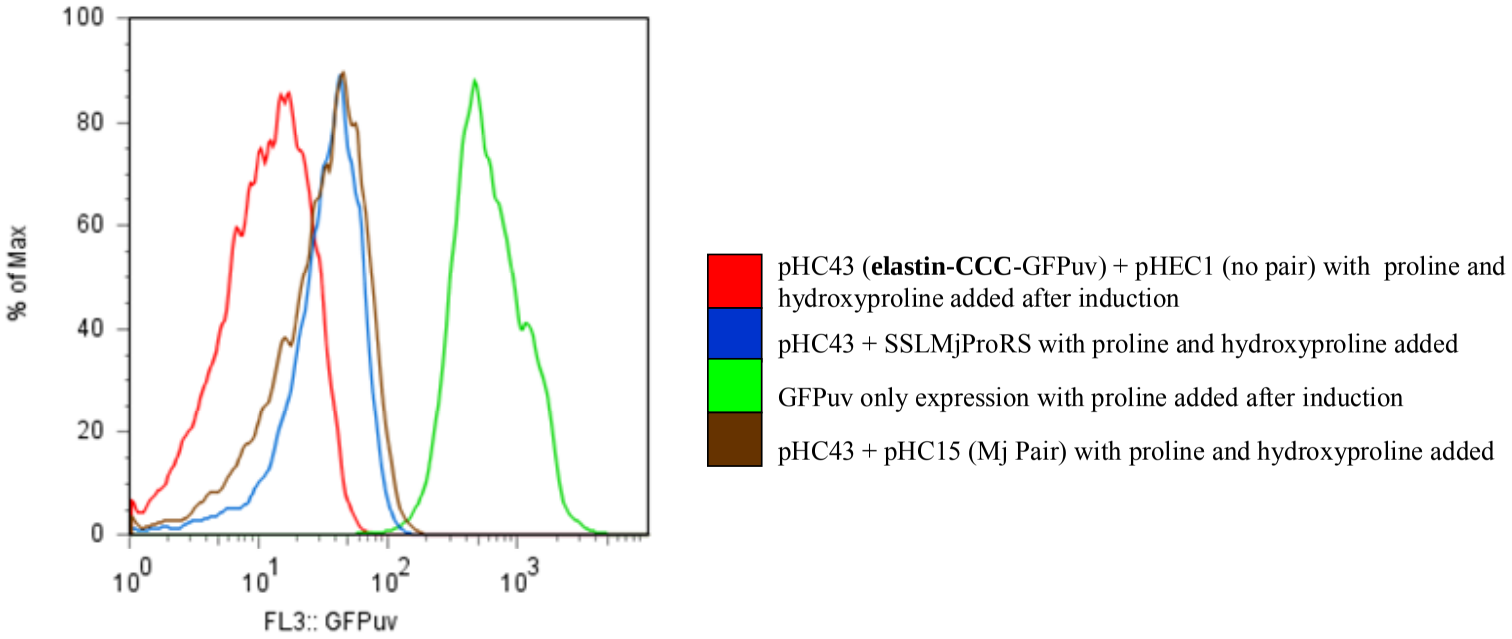


Figure 11. FACS analysis indicating the results of expression of the site-saturation library with the **elastin-CCC**-GFPuv reporter in the presence of the Mj pair compared to expression in the absence of the pair.

	Position 1	Position 2	Position 3	Position 4	Number of Mutations
Wild Type	GAA	TTT	GAG	TGC	
Mutant 1	GAA	TTT	GAG	TGC	None
Mutant 2	GAG (E103E)	ATG (F150M)	CAT (E152H)	CAG (C249E)	Three
Mutant 3	AG_ (Deletion)	CGG (F150R)	CCT (E152P)	AGT (C249S)	Four
Mutant 4	AAG (E103K)	TTT	GAG	TGC	One
Mutant 5	CTT (E103L)	GTT (F150V)	AAG (E152K)	TTG (C249L)	Four
Mutant 6	GGT (E103G)	TTT	GAG	AGG (C249R)	Two
Mutant 7	GTG (E103V)	TTT	GAG	TGC	One
Mutant 8	GAA	TTT	GAG	GGG (C249G)	One

Table 1. Results of MjProRS site-saturation library creation. Each of the four positions targeted for mutation (E103, F150, E152, and C249) were characterized by DNA sequencing to determine the diversity of the naïve library. The results above indicate that for mutants 1-8 picked at random from the library, mutations at all four sites or a subset of the sites were altered to produce a genetically diverse MjProRS library.

GFPuv and DsRed as potential reporter proteins

Other fluorescent reporters, besides GFPuv, were pursued later as targets for optimization of the FACS screening approach including DsRed-Express, GFPuv mutants, and superfolder GFP. For the purpose of directly screening mutant MjProRS enzymes with a preference for (2*S*,4*R*)-4-hydroxyproline, GFPuv and DsRed were investigated for use as reporter proteins in the absence of the elastin fusion. GFPuv was expressed in the presence of a variety of proline analogues (Figure 12). GFPuv was fluorescent when expressed with (2*S*)-proline, (2*S*,4*S*)-4-fluoroproline, (2*S*,3*R*)-3-fluoroproline, (2*S*,3*S*)-3-fluoroproline, (2*S*)-4,4-difluoroproline, (2*S*)-3,4-dehydroproline, (2*S*,4*S*)-4-hydroxyproline, and (4*R*)-1,3-thiazolidine-4-carboxylic acid. However, GFPuv was not fluorescent when expressed with (2*S*,4*R*)-4-fluoroproline, (2*S*,4*R*)-4-hydroxyproline, (2*S*)-piperidine-2-carboxylic acid, nor (2*S*)-azetidine-2-carboxylic acid. It was interesting to observe that the two analogues with a hydrophobic or hydrophilic functional group added at the C4 position with a *cis* conformation in the pyrrolidine ring produced a fluorescent protein while the same groups placed *trans* at the same position seemed to inhibit fluorescence upon incorporation. The conformations of proline analogues have been shown to play an important role in the stability of other proteins such as elastin-mimetic and collagen-mimetic proteins. Also, Budisa and coworkers recently found that the folding rates and stability of GFP are affected to a lesser extent by *cis/trans* isomerization of proline peptide bonds than by the puckering of pyrrolidine rings[26]. Upon global incorporation of (2*S*,4*R*)-4-fluoroproline and (2*S*,4*S*)-4-fluoroproline, the C^{γ}-*endo* conformation of the fluorine atoms stabilizes the GFP structure while the other

isomer has a less favorable effect. Interestingly, the incorporation of proline analogues with functional groups added at the C3 position did not seem to inhibit fluorescence.

Since (2*S*,4*R*)-4-hydroxyproline was a target for screening studies, the crystal structure of GFPuv (PDB code: 1B9C) was examined to determine a possible cause for the lack of fluorescence in GFPuv upon (2*S*,4*R*)-4-hydroxyproline incorporation. According to the crystal structure, 9 of the 10 proline residues in the crystal structure are in the *trans* conformation while one proline residue, P89, was found with a *cis* peptide bond conformation. Proline is normally found with a *trans* peptide bond conformation, so the *cis* peptide bond including P89 was believed to play an important role in the folding dynamics of GFPuv and a functional group, such as a fluorine atom or a hydroxyl group, in the *trans* conformation may alter folding dynamics of GFPuv resulting in a non-fluorescent protein. According to recent results from the Budisa group[26], the ring pucker of the proline analogue could be contributing to the folding and fluorescence of GFPuv.

In order to investigate the role of P89, a site-saturation library of mutants of the GFPuv gene was generated to produce a library of GFPuv proteins with all of the 20 common amino acids substituted for proline at position 89 in the protein sequence. While some colonies, expressing the GFPuv mutants, appeared to be fluorescent on plates after mutagenesis of the gene, when the fluorescent colonies were picked and the plasmid DNA within was sequenced, all of the clones encoded proline at the P89 position (Table 2). Several of the mutants from the naïve library had alternate proline codons indicating mutagenesis had worked, but proline was still encoded at the P89 position (Table 2). P89 is a critical residue for proper folding of GFPuv and may be an essential residue as no active mutants were found that did not have proline at the target position. No examples

of mutation at the P89 position were found in the literature. A site-mutant GFPuv was generated by site-directed mutagenesis where the proline at position 89 was converted to alanine. This mutant did not exhibit fluorescence upon expression.

	Color of colony	**P89 Genotype**	**Amino Acid Mutation**
Wild Type	Fluorescent Green	CCC	
Mutant 1	Fluorescent Green	CCC	None
Mutant 2	Fluorescent Green	CCG	None
Mutant 3	Fluorescent Green	CCC	None
Mutant 4	Fluorescent Green	CCG	None
Mutant 5	Fluorescent Green	CCG	None
Mutant 6	Fluorescent Green	CCC	None
Mutant 7	Fluorescent Green	CCC	None
Mutant 8	Fluorescent Green	CCC	None
Mutant 9	Fluorescent Green	CCT	None
Mutant 10	Fluorescent Green	CCC	None
Mutant 11	Fluorescent Green	CCT	None
Mutant 12	Fluorescent Green	CCC	None
Mutant 13	Fluorescent Green	CCC	None
Mutant 14	Fluorescent Green	CCC	None
Mutant 15	Fluorescent Green	CCG	None
Mutant 16	Fluorescent Green	CCC	None
Mutant 17	Fluorescent Green	CCC	None
Mutant 18	Fluorescent Green	CCC	None
Mutant 19	White	CGT	P89R
Mutant 20	White	AGG	P89R
Mutant 21	White	CTT	P89L
Mutant 22	White	TAG	P89Stop

Table 2. Results of site-saturation library creation in GFPuv. P89 was mutated to all possible amino acids using site-saturation mutagenesis. Fluorescent colonies were picked from a selective plate and DNA sequencing was used to screen for GFPuv mutants that were fluorescent or non-fluorescent.

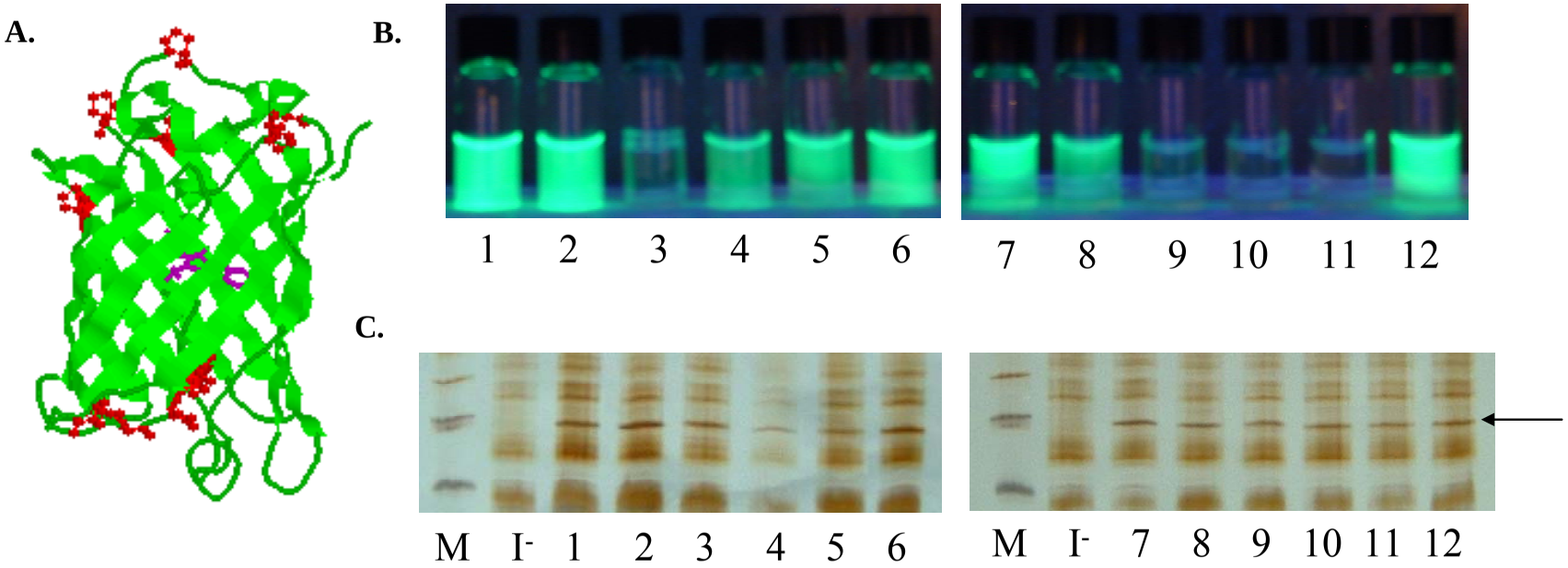


Figure 12. Fluorescence of purified GFPuv expressed with global incorporation of proline analogues. **A.** Structure of GFPuv indicating the location of proline residues in red. **B.** Fluorescence of purified samples of GFPuv after global proline analogue incorporation. **1.** (2*S*)-proline, **2.** (2*S*,4*S*)-4-fluoroproline, **3.** (2*S*,4*R*)-4-fluoroproline, **4.** (2*S*,3*R*)-3-fluoroproline, **5.** (2*S*,3*S*)-3-fluoroproline, **6.** (2*S*)-4,4-difluoroproline, **7.** (2*S*)-3,4-dehydroproline, **8.** (2*S*,4*S*)-4-hydroxyproline, **9.** (2*S*,4*R*)-4-hydroxyproline, **10.** (2*S*)-piperidine-2-carboxylic acid, **11.** (2*S*)-azetidine-2-carboxylic acid, and **12.** (4*R*)-1,3-thiazolidine-4-carboxylic acid. **C**. Results of expression of GFPuv with global incorporation of the analogues listed in **B**. above.

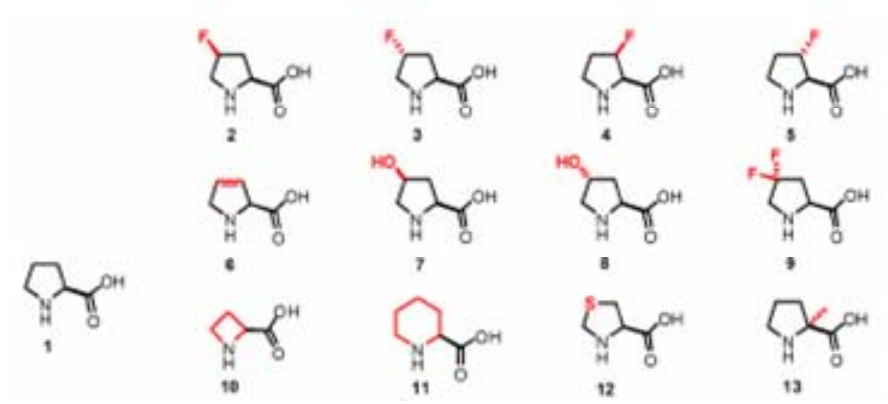


Figure 13. Chemical structures of proline (**1**) and the imino acid analogues (**2-13**): **1**. (2*S*)-proline; **2**. (2*S*,4*S*)-4-fluoroproline; **3**. (2*S*,4*R*)-4-fluoroproline; **4**. (2*R*,3*R*)-3-fluoroproline; **5**. (2*R*,3*S*)-3-fluoroproline; **6**. (2*S*)-3,4-dehydroproline; **7.** (2*S*,4*S*)-4-hydroxyproline; **8.** (2*S*,4*R*)-4-hydroxyproline; **9**. (2*S*)-4,4-difluoroproline; **10**. (2*S*)-azetidine-2-carboxylic acid; **11**. (2*S*)-piperidine-2-carboxylic acid; **12**. (4*R*)-1,3-thiazolidine-4-carboxylic acid; **13**, 2-methylproline

DsRed was then targeted for use as a fluorescent reporter. DsRed is a red fluorescent protein isolated from coral species of the *Discosoma* genus. DsRed-Express is a commercially available fluorescent reporter protein available from Clontech, Inc. DsRed-Express contains nine amino acid substitutions which enhance its solubility, reduce its green emission, and accelerate its maturation[27]. DsRed-Express exhibits a reduced tendency to aggregate as compared to the normally tetrameric wild-type DsRed protein. DsRed-Express contains 12 proline residues, all naturally encoded by the CCC codon, and the crystal structure of DsRed has been determined at a 2.0 Å resolution (PDB code: 1G7K). The overall structure of DsRed is very similar to GFP in that the DsRed monomer 3º structure is characterized by a β-barrel surrounding the mature chromophore and regions of random coil. DsRed would be useful as a reporter for the MjProRS synthetase activity screen as all of the proline residues are encoded by the codon of interest, CCC. Read-through of the DsRed protein would show activity of an orthogonal pair in the presence of hydroxyproline as long as hydroxyproline did not inhibit

fluorescence of DsRed. DsRed and GFPuv site-mutants were generated in an effort to generate active fluorescent variants with expression with hydroxyproline only.

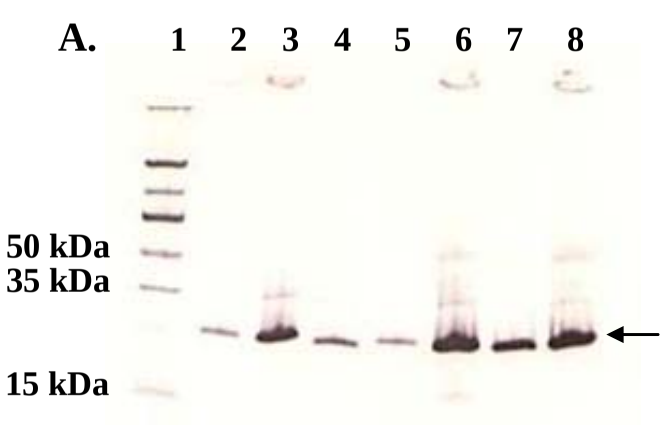
Use of GFPuv or DsRed alone (without the **elastin-CCC** fusion) would be desirable as the fluorescence levels were observed to be higher (visibly fluorescent) in cells expressing the reporter alone. Also, the DNA sequence of DsRed-Express indicated that there were 12 proline residues in the protein, and, interestingly, all of the prolines were encoded by the CCC codon. Expression of DsRed-Express with global incorporation of either (2*S*,4*R*)-4-hydroxyproline or (2*S*,4*S*)-4-hydroxyproline did not result in a fluorescent protein. When DsRed-Express was expressed with L-proline, the cells were brightly fluorescent, and western blotting analysis indicated a protein product of the correct size was produced in the case of proline or proline analogue incorporation (Fig. 14). The sequence indicated that there was a proline residue, P88, that was homologous to the P89 position in GFPuv. Also, upon examination of the crystal structure of DsRed (PDB code: 1G7K), one of the prolines, P63, was observed to lie directly adjacent to the residues forming the chromophore of DsRed. The P88 and P63 positions are likely important in controlling stability of the DsRed-Express protein. Directed evolution techniques could be a valuable approach for engineering a fluorescent reporter, for use in the MjProRS screening system, as the Tirrell group was able to generate a mutant GFP variant that exhibited fluorescence upon incorporation of 5,5,5-trifluoroleucine[28].

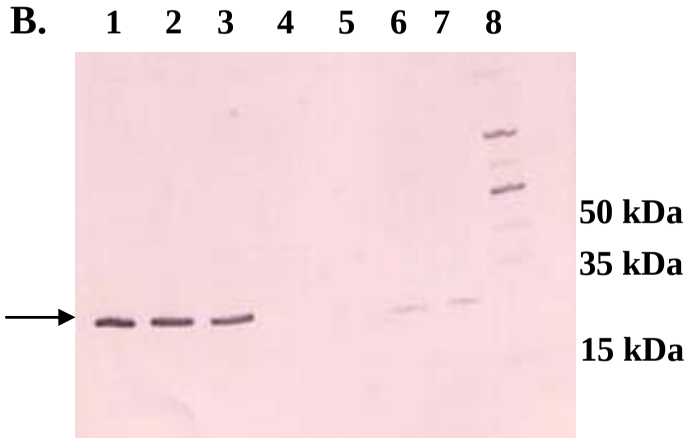
To investigate the role of the P63 position, a site-saturation library of DsRed-Express mutant proteins was generated where P63 was replaced with all 20 of the common amino acids. While some colonies, expressing the DsRed mutants, appeared to

be fluorescent (red) on plates after mutagenesis of the gene, when the fluorescent colonies were picked and the plasmid DNA within was isolated and sequenced, all of the clones encoded proline at the P63 position. Several of the mutants had alternate proline codons indicating mutagenesis had worked, but proline was still encoded at the P63 position (Table 3). One non-fluorescent (white) and one partially fluorescent (pink) colony were picked and DNA sequencing indicated that the P63 position had been changed to serine in the white colony and changed to cysteine in the pink colony. It appeared that the P63 position was important for folding and fluorescence of DsRed-Express, but that low levels of fluorescence compared to wild-type could be achieved by mutation to alternate residues (P63C).

	Color of colony	**P63 Genotype**	**Amino Acid Mutation**
Wild Type	Red	CCC	
Mutant 1	Red	CCC	None
Mutant 2	Red	CCC	None
Mutant 3	Red	CCC	None
Mutant 4	Red	CCC	None
Mutant 5	Red	CCC	None
Mutant 6	Red	CCG	None
Mutant 7	White	AGT	P63S
Mutant 8	Pink	TGT	P63C

Table 3. A site-saturation library of DsRed mutants was constructed in which the P63 position was mutated to all possible amino acids. Six red colonies, one white colony, and one pink colony were picked at random (containing the naïve library), the DNA was isolated from the bacterial host, and the DsRed sequence was determined by DNA sequencing.





1. Perfect protein marker

2. pHC47 (DsRed) + pHEC1 (no pair), no proline

3. pHC47 + pHEC1, proline added after induction

4. pHC47 + pHEC1, *cis*-4-hydroxyproline

5. pHC47 + pHEC1, *trans*-4-hydroxyproline

6. pHC47 + pHC15 (Mj pair), proline added

7. pHC47 + pHC15, *cis*-4-hydroxyproline

8. pHC47 + pHC15, *trans*-4-hydroxyproline

1. pHC47 (DsRed) + pHC15, *trans*-4-fluoroproline

2. pHC47 + pHC15, *cis*-4-fluoroproline

3. pHC47 + pHC15 (Mj pair), proline added

4. pHC47 + pHEC1, *trans*-4-fluoroproline

5. pHC47 + pHEC1, *cis*-4-fluoroproline

6. pHC47 + pHEC1, proline added after induction

7. pHC47 + pHC15, no proline added

8. Perfect protein marker

Figure 14. Western blot analysis of expressions of DsRed with proline analogue incorporation. Full-length DsRed is produced (arrows) but analogue incorporation seems to inhibit fluorescence of the protein.

Elastin-CCC-superfolder GFPuv fusion reporter

The superfolder GFP protein was of particular interest for use in development of a screen in that the variant of GFP was subjected to mutagenesis to generate a fluorescent protein with enhanced folding robustness[23]. Directed evolution and FACS-based techniques were used to identify GFP variants that folded properly even when fused to poorly folded proteins or inclusion bodies[23]. The resultant superfolder GFP had six new amino acid mutations that were shown to be responsible for the enhanced folding capability (S30R, Y39N, N105T, Y145F, I171V, and A206V)[23]. Superfolder GFP could potentially be used as a fusion to **elastin-CCC** or alone in screening for MjProRS synthetase variants that recognized hydroxyproline over proline as the folding and fluorescence properties of superfolder would be enhanced compared to GFPuv even when fused to the poorly folded **elastin-CCC** protein.

The **elastin-CCC**-superfolder GFP reporter was of interest as a reporter system as the GFP variant protein was optimized, using directed evolution techniques, to fold well even when fused to poorly folded proteins or inclusion bodies[23]. **Elastin-CCC** is insoluble at the *E. coli* growth temperature used for expression (37 ºC) and it was suspected that misfolding or aggregation of **elastin-CCC** was interfering with the proper folding of GFPuv in the fusion reporter system. A C-terminal **elastin-CCC**-superfolder gene cassette was designed for use as a reporter in screening of MjProRS mutant libraries (Fig. 15). Since enrichment of the MjProRS mutant library in clones selective for hydroxyproline was not successful with the **elastin-CCC**-GFPuv fusion reporter, the **elastin-CCC**-superfolder GFP reporter construct was a better alternative as fluorescence levels of the reporter were higher. If fluorescence is too low, the method is not effective

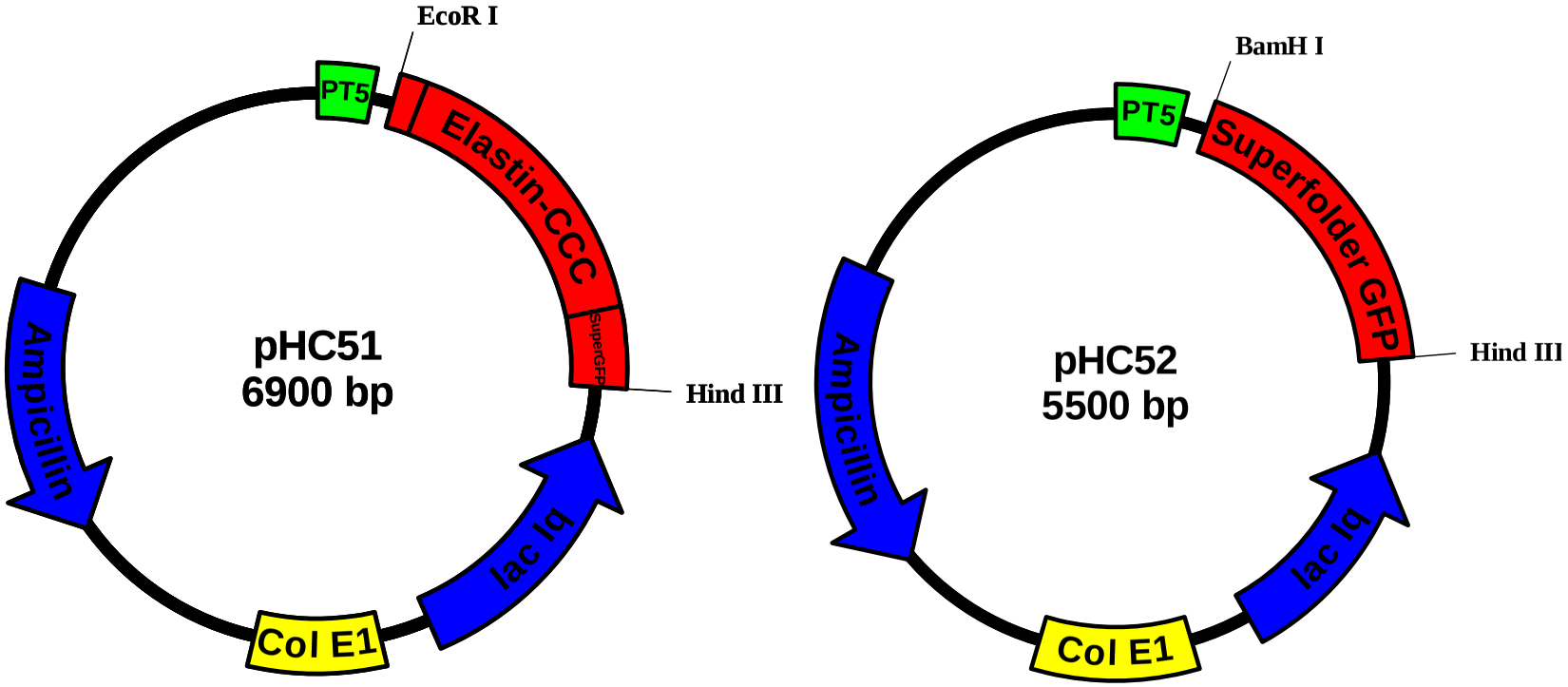


Figure 15. Plasmids for the expression of fusions of **elastin-CCC** to superfolder GFP. The pHC51 plasmid was generated by cloning the superfolder GFP gene into the pQE-80L expression plasmid harboring the **elastin-CCC** gene cassette to generate the C-terminal **elastin-CCC**-superfolder GFP fusion reporter. The pHC52 plasmid was generated by cloning the superfolder GFP gene into the pQE-80L expression plasmid.

in sorting negative vs. positively fluorescent cells. The MjProRS site-saturation library was transformed into cells containing the expression plasmid harboring the **elastin-CCC**-superfolder GFP gene. FACS analysis revealed promising differences between the negative (no pair present) and positive (Mj pair co-expressed) controls as well as the potential for sorting the library with the new reporter (Fig. 16). The **elastin-CCC**-superfolder GFP produced a much higher fluorescence signal compared to expressions with the **elastin-CCC**-GFPuv reporter in the case of the fusion as well as with the fluorescent reporter alone (Figs. 16 and 17). In Figure 16, expression of the fluorescent reporter protein was carried out with co-expression of the site-saturation library of MjProRS variants. The negative control (red) indicated very low to no fluorescence when proline was not added to the media with expression of the reporter in the proline auxotrophic strain, UQ27ts, ProA. The starting level of cellular fluorescence (blue) was analyzed and a gate was drawn around the most highly fluorescent cells. Sorting one time in purity mode allowed for enrichment of highly fluorescent cells (brown). Superfolder GFP expression alone is shown in green. Expression of superfolder GFP either alone or as a C-terminal fusion to **elastin-CCC** demonstrated a much higher overall cellular fluorescence compared to the GFPuv reporter systems. In Figure 17, fluorescence of cells with L-proline incorporation in superfolder GFP was slightly higher than that with hydroxyproline incorporation. GFPuv and DsRed were not fluorescent when expressed with global incorporation of (2*S*,4*R*)-4-hydroxyproline. The stability and/or folding dynamics of superfolder GFP may allow for proper folding and fluorescence of the protein even in the presence of hydroxyproline. Fluorescence of cells before induction of protein expression is shown in red.

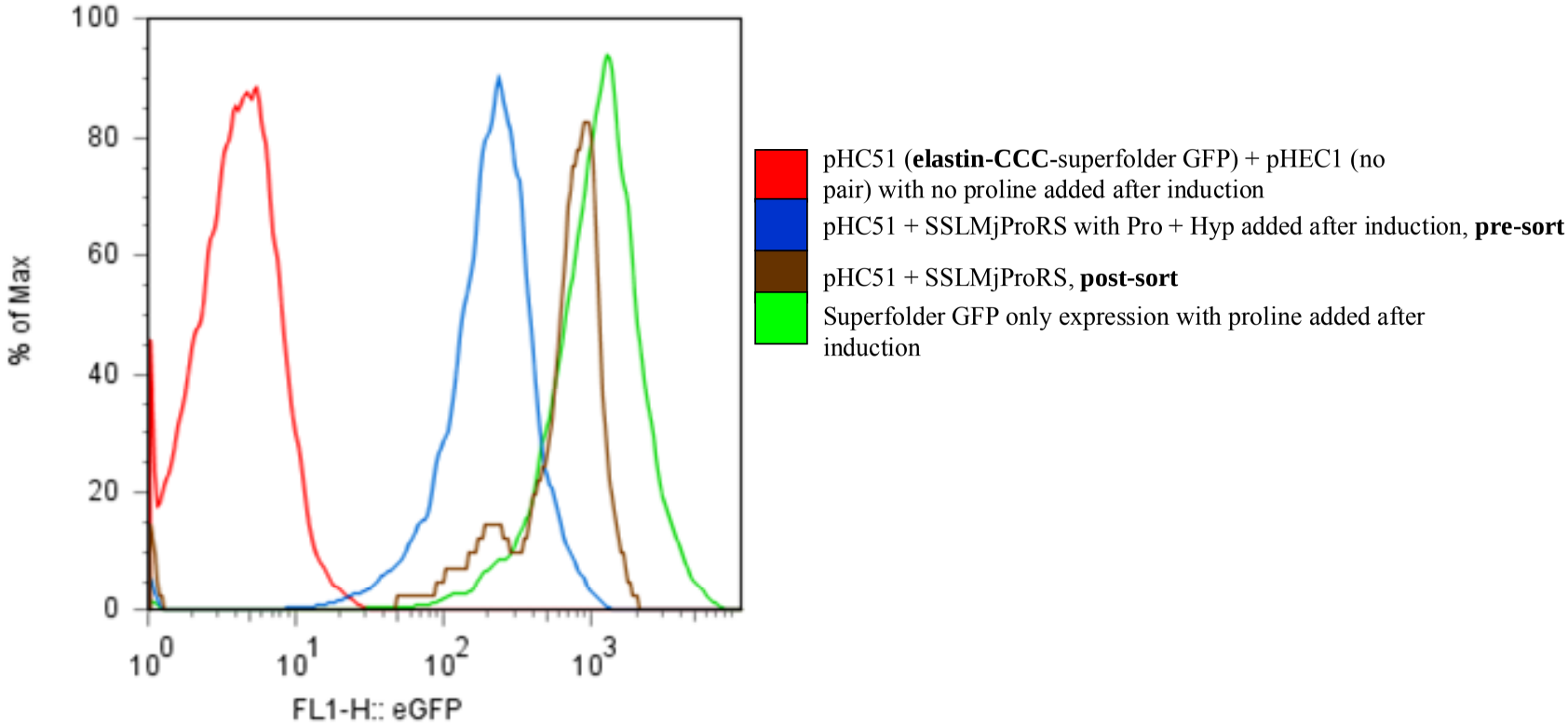


Figure 16. Example of one round of positive sorting of the MjProRS site-saturation mutant library with FACS using the fluorescent reporter, **elastin-CCC**-superfolder GFP.

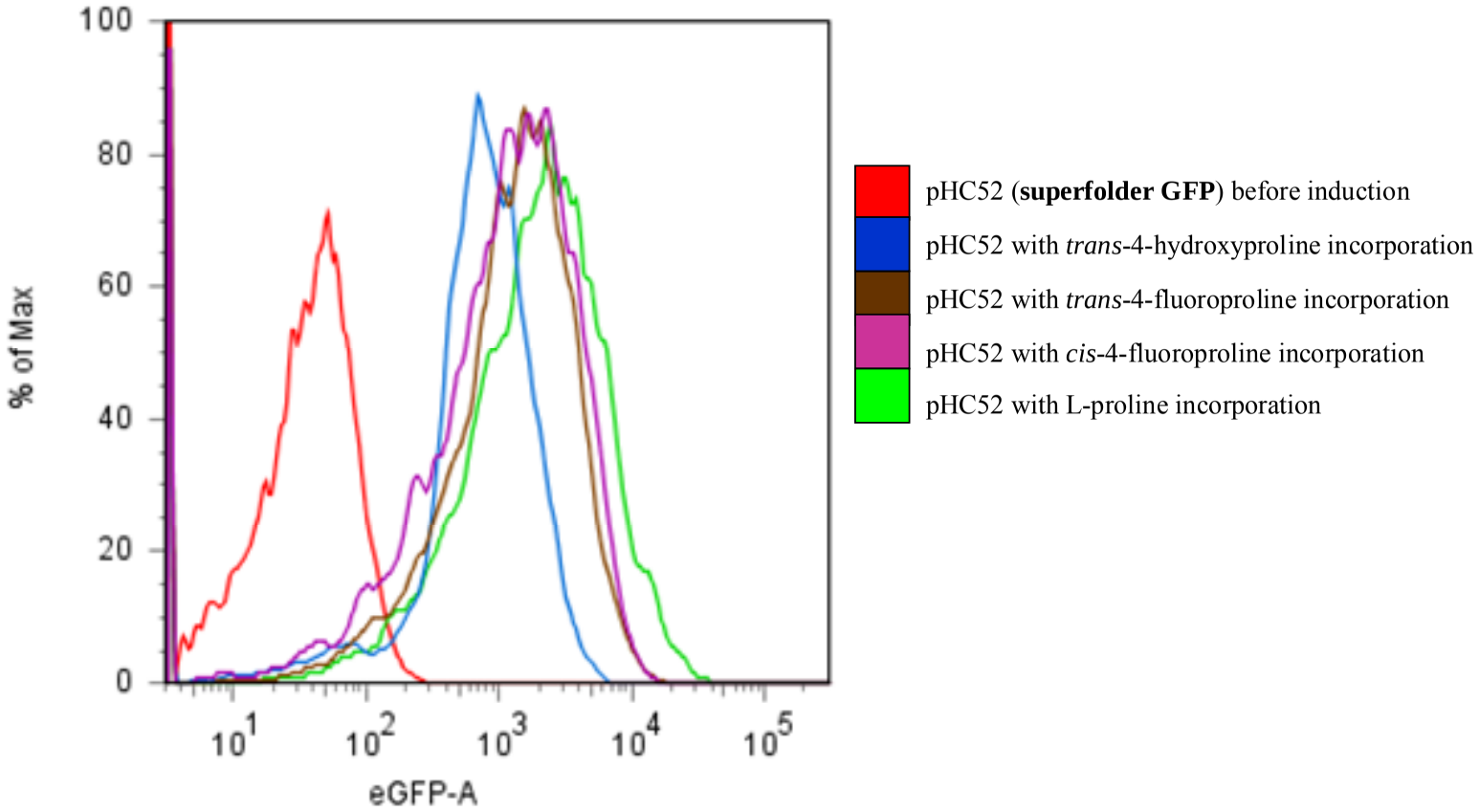


Figure 17. FACS analysis of fluorescent cells expressing superfolder GFP (alone) with global incorporation of proline versus global incorporation of proline analogues. Expression of the fluorescent reporter protein indicated that the protein was fluorescent with L-proline (green) and with (2*S*,4*R*)-4-hydroxyproline (>99 % pure) incorporation (blue) at 37 ºC as well as with the fluorinated analogues.

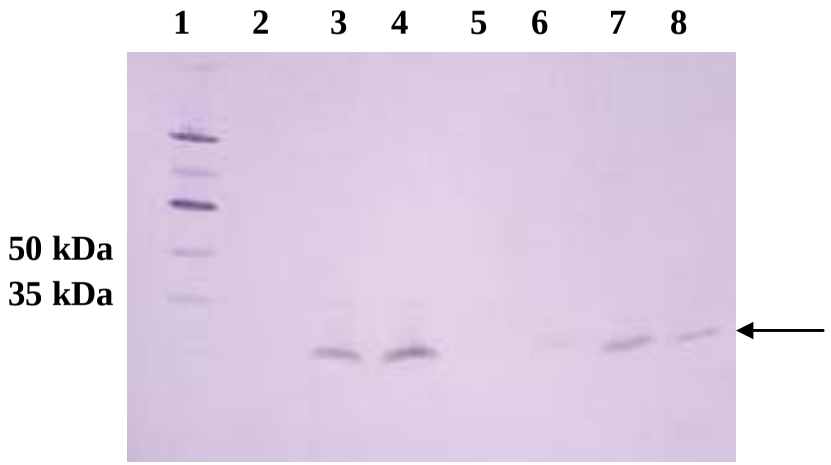
Another advantage of using the **elastin-CCC**-superfolder GFP reporter is the ability to conduct expressions for screening the mutant library at 42 ºC rather than at 37 ºC. With mutation of all of the proline codons (10 CCG codons) in the superfolder GFP sequence to the CCC codon, expressions with proline analogues Using the GFPuv reporter, the UQ27$^{ts,\ ProA}$ strain is grown at 30 ºC initially until the cells reach log phase growth, then protein expression is induced at 37 ºC to knock down expression of the endogenous EcProRS in the mutant strain. Knocking down expression of the endogenous levels of EcProRS reduced the background levels of fluorescence resultant from activity of the EcProRS in decoding the CCC codons of **elastin-CCC**. Some amount of the EctRNA^{Pro} is still present even at very low levels in the cells which would lead to decoding of the CCC codons if the EcProRS is active. If background fluorescence is too high, sorting between positive and negative cells is difficult as there is not a big difference between the cells that are "off" and those that are highly fluorescent. The **elastin-CCC**-superfolder GFP protein is fluorescent if expressed at 30 ºC or 37 ºC in the UQ27$^{ts,\ ProA}$ strain. Screening the MjProRS mutant library at 42 ºC would be ideal as activity of the EcProRS is dramatically reduced and very low background level fluorescence would likely result in larger differences between the negative control and cells expressing the reporter at a higher level due to activity of a mutant synthetase specific for hydroxyproline.

Another significant advantage to using superfolder GFP over GFPuv is that superfolder GFP was shown to be fluorescent with global incorporation of (2*S*,4*R*)-4-hydroxyproline as well as with fluorinated proline analogues (Fig. 17). Both GFPuv and DsRed were non-fluorescent when expressed with global incorporation of (2*S*,4*R*)-4-

hydroxyproline. Superfolder GFP could be engineered to be used alone as a reporter for screening MjProRS synthetase mutant activity toward (2*S*,4*R*)-4-hydroxyproline. By engineering silent mutations at each of the proline codons (10 sites) in superfolder GFP, all of the codons could be converted to CCC codons that could be used to show read-through of the reporter gene without the **elastin-CCC** fusion protein. The MjProRS mutants that were specific for (2*S*,4*R*)-4-hydroxyproline would charge the CCC sites with the proline analogue at specific positions and produce cells exhibiting a higher fluorescence level than those expressing the wild type MjProRS. Since the EcProRS would be inactive when cells were induced at 42 ºC, the background level of fluorescence would be expected to be very low.

The superfolder GFP protein was over-expressed in *E. coli* with proline and (2*S*,4*R*)-4-hydroxyproline incorporation on a 500 mL scale for each expression. The proteins were harvested from bacterial cells for characterization using metal affinity chromatography. The hexahistidine tags at the N and C termini of the superfolder GFP protein were useful in the purification of the material. Western bolt analysis revealed the level of expression was lower for superfolder GFP with hydroxyproline incorporation compared to proline incorporation (Fig. 18). Incorporation of the fluorinated analogues, (2*S*,4*R*)-4-fluoroproline and (2*S*,4*S*)-4-fluoroproline also resulted in a fluorescent superfolder GFP with an intermediate expression level compared to that of proline and hydroxyproline incorporation levels (Fig. 18). SDS-PAGE analysis was used to evaluate the purity of the superfolder GFP proteins in the case of proline incorporation compared to hydroxyproline incorporation (Fig. 19). No detectable superfolder GFP was present in the pellet fraction with all of the soluble protein in the supernatant fraction after protein

expression. Purification of the proteins via metal affinity chromatography resulted in a yield of 15 mg superfolder GFP with hydroxyproline incorporation compared to 50 mg superfolder GFP with proline incorporation with both samples at high purity after dialysis (Fig. 19). Fourier-transform mass spectrometry (FTMS) with electrospray ionization was used to determine the relative purity of each sample as well (Fig. 20). The technique provided higher resolution compared to MALDI-TOF-MS analysis as MALDI resulted in single large peaks in the case of proline incorporation in superfolder GFP as well as in the case of hydroxyproline incorporation. With proline incorporation, the predominant species in the sample had a m/z ratio at 29,256.7 and with hydroxyproline incorporation there were two predominant species with m/z ratios at 29,239.7 and 29,403.6. It appeared that there were several mixed species expressed in the case of hydroxyproline incorporation which could be explained by residual proline in the system or a proline salvage pathway in the organism resulting in incorporation of some proline and hydroxyproline in superfolder GFP. The species at the m/z ratio of 29,239.7 likely represented superfolder GFP with proline incorporation, and the species at 29,403.6 had a m/z ratio representative of hydroxyproline incorporation with a mass difference of ~160 Da. The theoretical mass calculated for superfolder GFP with proline incorporation, based on the amino acid sequence of the protein with two hexahistidine tags was 29,002 g/mol and the mass calculated for superfolder GPF with hydroxyproline incorporation was 29,162 g/mol. The difference in the theoretical masses and the m/z ratios observed may have been a result of Na^+ adherence to the superfolder protein even after dialysis in pure water.



1. Perfect protein marker

2. pHC52 (superfolder GFP), with no proline added

3. pHC52 with proline added

4. pHC52 with proline added (large scale expression)

5. pHC52 with *trans*-4-hydroxyproline added

6. pHC52 with hydroxyproline (large scale expression)

7. pHC52 with *trans*-4-fluoroproline added

8. pHC52 with *cis*-4-fluoroproline added

Figure 18. Western blot analysis of expressions of superfolder GFP with incorporation of proline and proline analogues.

1. Perfect protein marker

2. pHC52 (superfolder GFP), with proline, pellet

3. pHC52 with hydroxyproline, pellet

4. pHC52 with proline, supernatant

5. pHC52 with hydroxyproline, supernatant

6. pHC52 with proline, before dialysis

7. pHC52 with proline, after dialysis (pure)

8. pHC52 with hydroxyproline, after dialysis (pure)

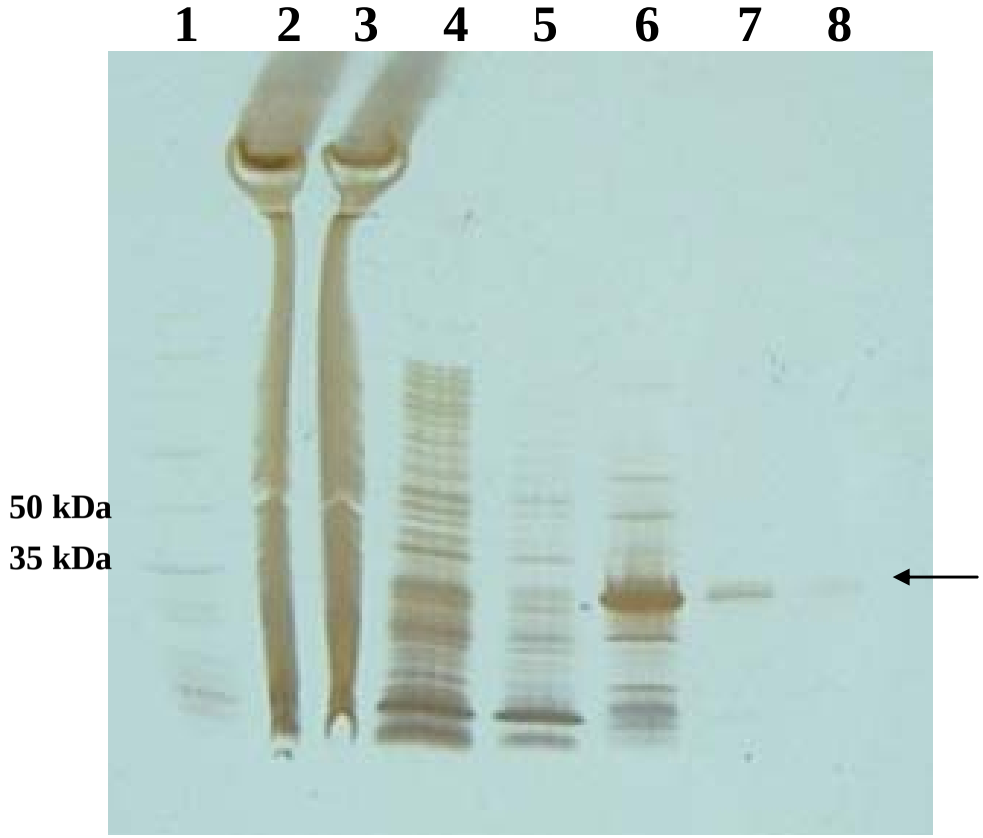


Figure 19. SDS-PAGE analysis indicating purification of superfolder GFP with incorporation of proline and (2*S*,4*R*)-4-hydroxyproline.

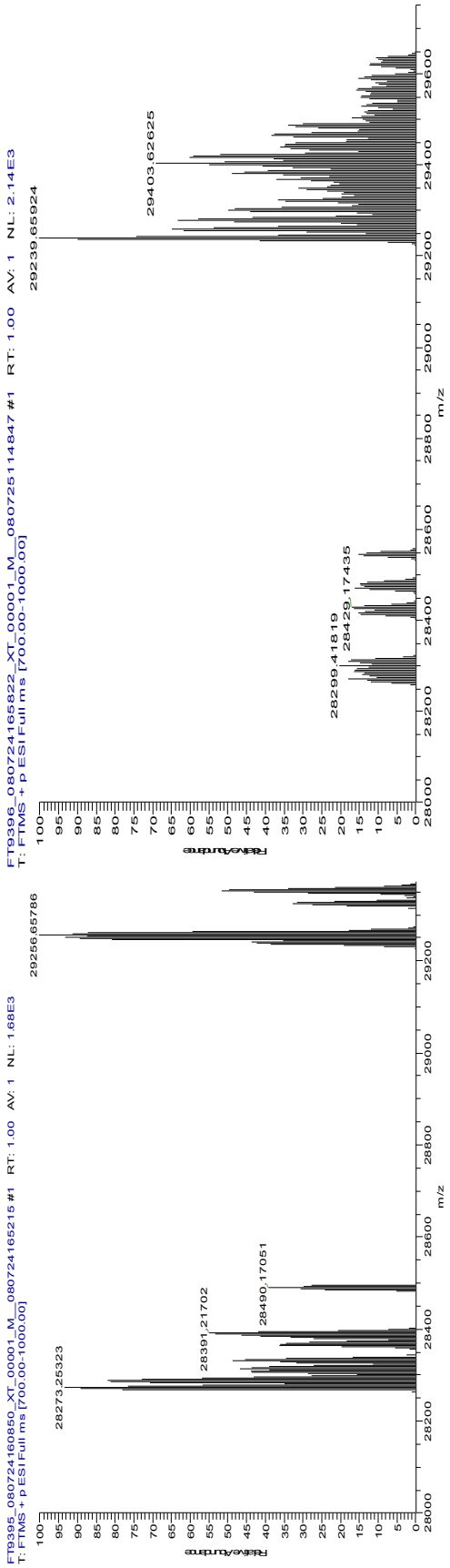


Figure 20. FT-MS with electrospray ionization indicating the m/z ratio of various species isolated from superfolder GFP expression and purification. A. Results indicating m/z ratios for species isolated from expression of superfolder GFP with proline incorporation. B. Results indicating m/z ratios for species isolated from expression of superfolder GFP with *trans*-4-hydroxyproline incorporation.

A controlled evaluation of the FACS sorting technique was carried out to determine the effectiveness of the method in the mutant MjProRS screening strategy. Expression of the fluorescent reporter protein, GFPuv, was carried out at 37 ºC. The **elastin-CCC**-GFPuv reporter was also expressed at 37 ºC, with co-expression of an empty plasmid (non-fluorescent) or with the Mj orthogonal pair (fluorescent). Mixtures of fluorescent cells and non-fluorescent cells were prepared at 10:1 volume ratios (10 non-fluorescent: 1 fluorescent), and FACS analysis results can be seen in Figure 21. In section **A**. with the fluorescent reporter (GFPuv) alone, the two populations (fluorescent and non-fluorescent) of cells in a single sample are easily detected with FACS analysis using the side-scatter parameter (arrows). In **B.** of Fig. 21, the **elastin-CCC**-GFPuv reporter is not highly fluorescent and cells expressing the protein were not bright enough to detect using the side scatter parameter at a 10:1 ratio. These results indicate that a brighter fluorescent reporter, such as superfolder GFP is more favorable for FACS-based sorting techniques as a weak reporter may fall under the technique's limit of detection with prokaryotic cells.

A control sort using the GFPuv reporter was carried out to determine its effectiveness in screening methodologies compared to superfolder GFP. Expression of the fluorescent reporter protein, GFPuv, was carried out at 37 ºC. Non-fluorescent cells were prepared by inducing expression in cells which did not contain the reporter gene (empty plasmid). Mixtures of fluorescent cells and non-fluorescent cells were prepared at 10:1, 100:1, and 1000:1 volume ratios (10 or 100 or 1000 non-fluorescent: 1 fluorescent). As seen in FACS analysis in Figure 22 (**A.)**, a decrease in detection of fluorescent cells (arrow) is observed with increasing bias in the identity (fluorescence

intensity) of cells in mixed cell populations. In **B.,** FACS sorting of cells expressing the fluorescent GFPuv reporter alone was carried out to determine if fluorescent cells could be sorted from a mixed population. The 1000:1 mixed population was sorted for the most highly fluorescent cells. A gate in the fluorescence channel was drawn around the most highly fluorescent cells (0.5 %) of the mixed population. FACS analysis (arrow) indicated an enrichment after sorting where highly fluorescent cells represented ~14 % of the sorted population of cells.

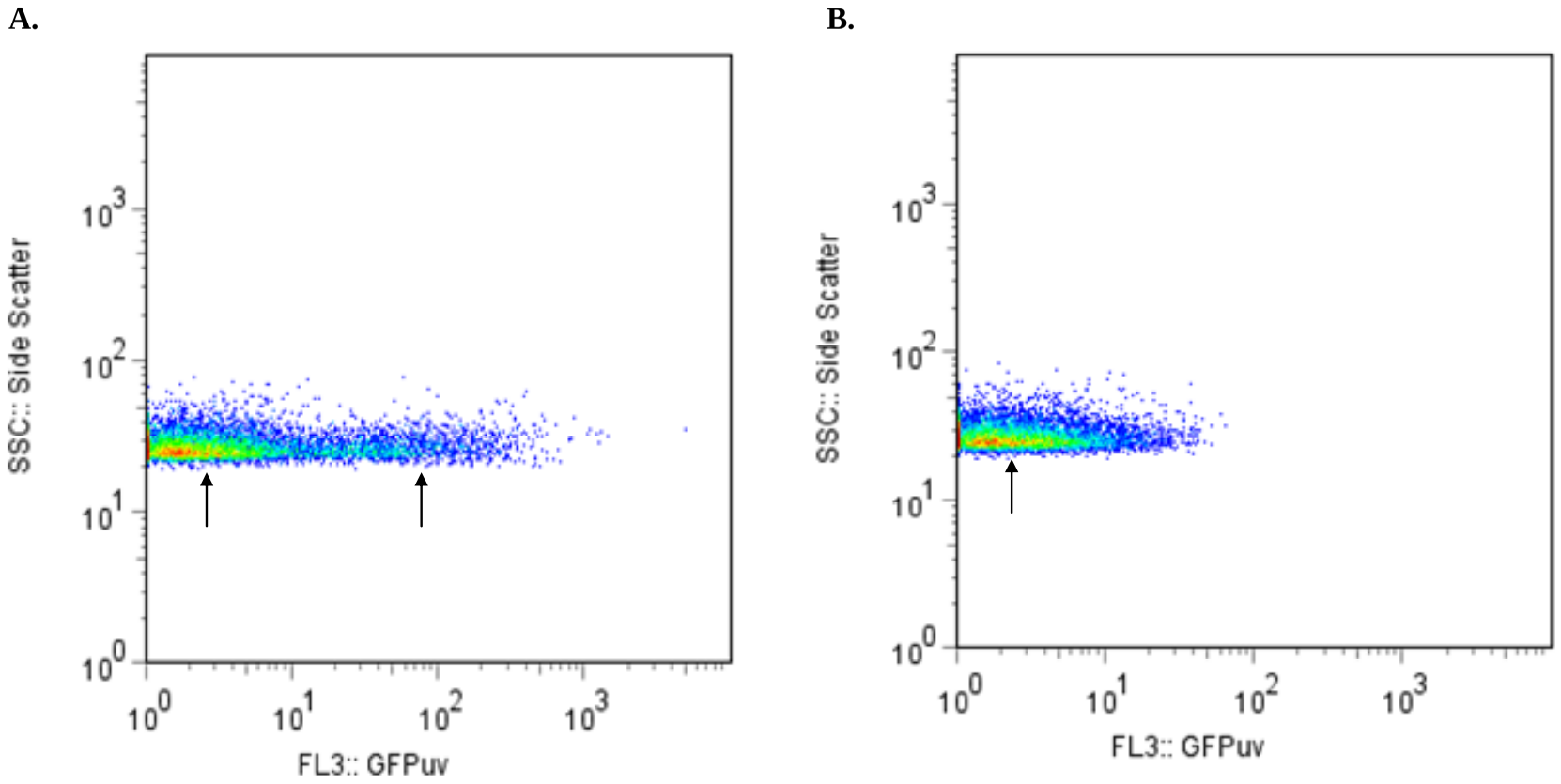


Figure 21. Controlled evaluation of the FACS sorting technique. Expression of the fluorescent reporter protein, GFPuv, was carried out at 37 ºC. The **elastin-CCC**-GFPuv reporter was also expressed at 37 ºC, with co-expression of an empty plasmid (non-fluorescent) or with the Mj orthogonal pair (fluorescent). Mixtures of fluorescent cells and non-fluorescent cells were prepared at 10:1 volume ratios (10 non-fluorescent: 1 fluorescent).

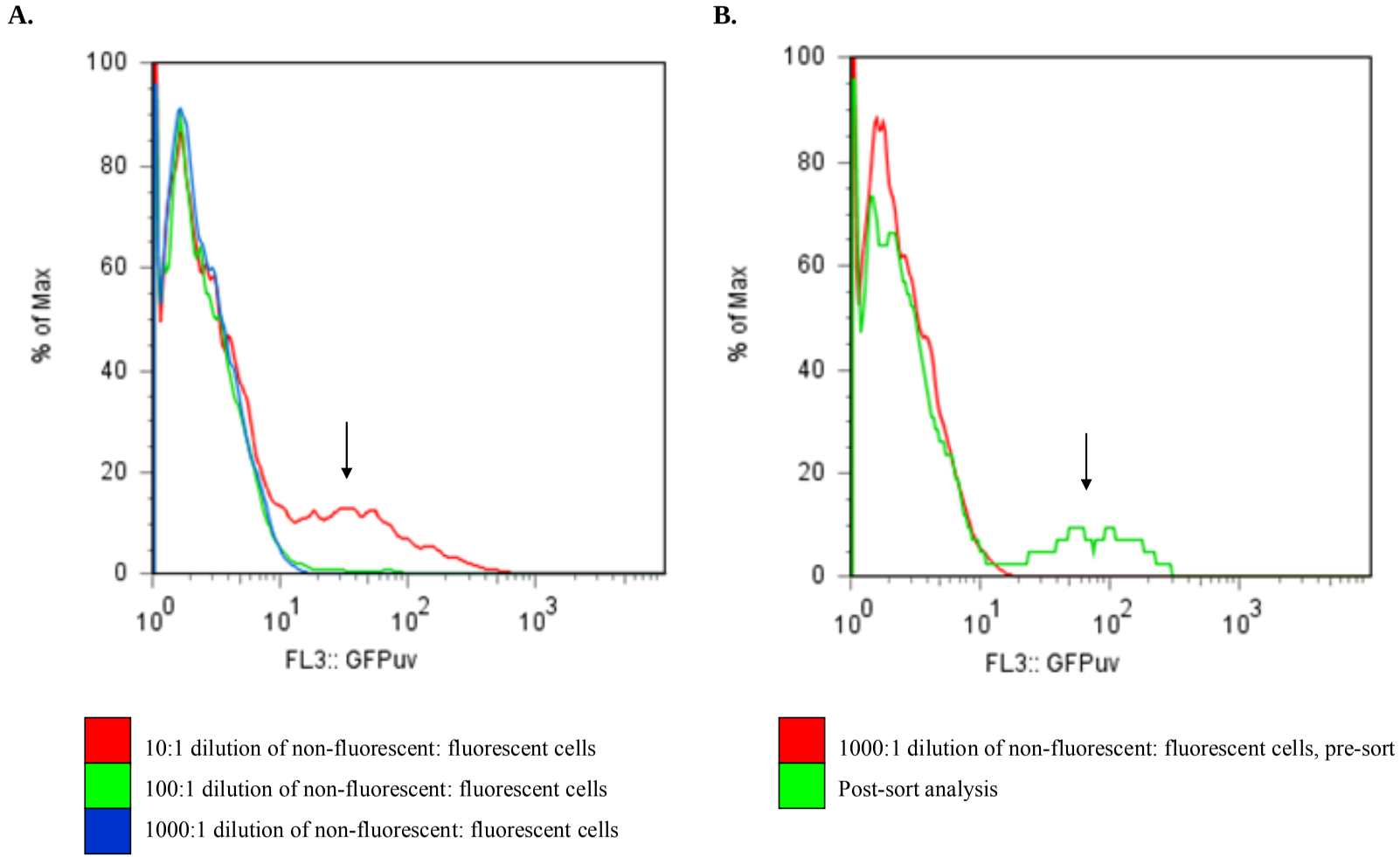


Figure 22. Control sort using the GFPuv reporter. Expression of the fluorescent reporter protein, GFPuv, was carried out at 37 ºC. Non-fluorescent cells were prepared by inducing expression in cells which did not contain the reporter gene (empty plasmid). Mixtures of fluorescent cells and non-fluorescent cells were prepared at 10:1, 100:1, and 1000:1 volume ratios (10 or 100 or 1000 non-fluorescent: 1 fluorescent).

1. pAG2 + pHC15 (Mj pair), hydroxyproline added

2. pAG2 (**elastin-1**) + pHC15, proline added

3. pAG2 + pHEC1 (no pair), hydroxyproline added

4. pAG2 + pHEC1, proline added

5. pHC8 (**elastin-CCC**) + pHC15, hydroxyproline added

6. pHC8 + pHC15, proline added

7. pHC8 + pHEC1, proline incorporation

8. Perfect protein marker

Figure 23. SDS-PAGE analysis indicating expression of **elastin-CCC** or **elastin-1** at 42 ºC in the UQ27^{ts,ProA} strain with incorporation of proline and (2*S*,4*R*)-4-hydroxyproline. Expression of **elastin-1** was not possible in the presence or absence of the Mj pair with hydroxyproline incorporation. However, expression of **elastin-CCC** was possible with proline and hydroxyproline incorporation due to the activity of the Mj pair.

Conclusions

Rational design of the MjProRS enzyme with alteration of 1-2 specific amino acid residues by site-directed mutagenesis did not result in a significant change in specificity of the MjProRS enzyme. Random and site-saturation mutagenesis were chosen as alternate methodologies since rational design did not yield promising results after several trials. Random mutagenesis was not pursued at great length as a method for altering substrate specificity in the MjProRS enzyme, but the method was optimized for future library construction and may be a viable option for refinement of a library in the future. In addition, the random mutagenesis technique was not pursued as site-saturation mutagenesis was shown to be successful in the literature. After discussion with many members of the directed evolution field, it was determined that there is, currently, not a single formula method for conducting directed evolution of enzymes[14]. Many times, a combination of several methods are used to alter the substrate specificity of an enzyme including random mutagenesis, site-saturation mutagenesis, DNA shuffling, and circular permutation among other methods[14,29]. Site-saturation mutagenesis was pursued as a more specific way to alter the specificity of the MjProRS enzyme with the goal of finding an enzyme that showed a preference for (2*S*,4*R*)-4-hydroxyproline. Random mutagenesis or DNA shuffling techniques may be used in the future, in addition to site-saturation mutagenesis, in engineering synthetase enzymes.

Site-saturation mutagenesis was pursued aggressively as a method for altering substrate specificity but several rounds of sorting did not yield a mutant enzyme with the desired hydroxyproline specificity using the **elastin-CCC**-GFPuv fusion reporter. Perhaps other sites, either other residues in the active site or residues outside the active

site, must me chosen for mutagenesis to maximize potential for isolation of the desired enzyme activity. Development of an alternate reporter screen is promising as the superfolder GFP reporter is both more highly fluorescent than the GFPuv fusion system previously used and the superfolder GFP reporter is active with (2*S*,4*R*)-4-hydroxyproline only incorporation. GFPuv, however, was not active with (2*S*,4*R*)-4-hydroxyproline only incorporation.

It may be necessary to generate a knockout version of the UQ27$^{ts,\ ProA}$ strain in which the tRNA used for decoding the CCC codon (EctRNA^{ProL}) is not produced in the expression strain for production of protein materials with simultaneous proline and hydroxyproline incorporation. An *in vitro* assay may also be useful for characterization of the kinetics of binding and activity of the MjProRS enzyme toward proline and hydroxyproline as a method for screening MjProRS mutant individuals that show promise in fluorescence (FACS)-based or western blot/expression-based assays for activity. Overall, the site-saturation mutagenesis technique was optimized for use in directed evolution of MjProRS variants and a screening system was developed for FACS-based isolation of MjProRS mutant enzymes with activity for (2*S*,4*R*)-4-hydroxyproline.

References

(1) Farinas, E. T.; Bulter, T.; Arnold, F. H. *Current Opinion in Biotechnology* **2001**, *12*, 545-551.

(2) Bloom, J. D.; Meyer, M. M.; Meinhold, P.; Otey, C. R.; MacMillan, D.; Arnold, F. H. *Current Opinion in Structural Biology* **2005**, *15*, 447-452.

(3) Ryu, Y. H.; Schultz, P. G. *Nature Methods* **2006**, *3*, 263-265.

(4) Wang, L.; Xie, J.; Schultz, P. G. *Annual Review of Biophysics and Biomolecular Structure* **2006**, *35*, 225-249.

(5) Chen, S.; Schultz, P. G.; Brock, A. *Journal of Molecular Biology* **2007**, *371*, 112-122.

(6) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

(7) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Journal of the American Chemical Society* **2002**, *124*, 9026-9027.

(8) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nature Biotechnology* **2002**, *20*, 1044-1048.

(9) Shapiro, H. M. *Practical Flow Cytometry*; 4th ed.; John Wiley and Sons, Inc.: Hoboken, NJ, 2003.

(10) Garrett, R. H., Grisham, C.M. *Biochemistry*; 2nd ed.; Thomson, 1999.

(11) Yoo, T. H.; Tirrell, D. A. *Angewandte Chemie-International Edition* **2007**, *46*, 5340-5343.

(12) Stehlin, C.; Heacock, D. H.; Liu, H.; Musier-Forsyth, K. *Biochemistry* **1997**, *36*, 2932-2938.

(13) Kim, W. Y.; George, A.; Evans, M.; Conticello, V. P. *Chembiochem* **2004**, *5*, 928-936.

(14) Arnold, F. H. *Nat Biotech* **2006**, *24*, 328-330.

(15) Li, Y.; Drummond, D. A.; Sawayama, A. M.; Snow, C. D.; Bloom, J. D.; Arnold, F. H. *Nat Biotech* **2007**, *25*, 1051-1056.

(16) Wang, L.; Schultz, P. G. *Chemical Communications* **2002**, 1-11.

(17) Buechter, D. D.; Paolella, D. N.; Leslie, B. S.; Brown, M. S.; Mehos, K. A.; Gruskin, E. A. *Journal of Biological Chemistry* **2003**, *278*, 645-650.

(18) Beuning, P. J.; Musier-Forsyth, K. *Journal of Biological Chemistry* **2001**, *276*, 30779-30785.

(19) Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C. *Nat Biotech* **1996**, *14*, 315-319.

(20) Singer, M.; Baker, T. A.; Schnitzler, G.; Deischel, S. M.; Goel, M.; Dove, W.; Jaacks, K. J.; Grossman, A. D.; Erickson, J. W.; Gross, C. A. *Microbiological Reviews* **1989**, *53*, 1-24.

(21) Bohman, K.; Isaksson, L. A. *Molecular & General Genetics* **1980**, *177*, 603-605.

(22) Gabriel, K.; McClain, W. H. *Journal of Molecular Biology* **2001**, *310*, 543-548.

(23) Pedelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat Biotech* **2006**, *24*, 79-88.

(24) Studier, F. W. *Protein Expression and Purification* **2005**, *41*, 207-234.

(25) Sambrook, J.; Russell, D. W. *Molecular Cloning: A laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

(26) Steiner, T.; Hess, P.; Bae, J. H.; Wiltschi, B.; Moroder, L.; Budisa, N. *PLoS ONE* **2008**, *3*, e1680.

(27) Bevis, B. J.; Glick, B. S. *Nat Biotech* **2002**, *20*, 83-87.

(28) Yoo, T. H.; Link, A. J.; Tirrell, D. A. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 13887-13890.

(29) Park, H.-S.; Nam, S.-H.; Lee, J. K.; Yoon, C. N.; Mannervik, B.; Benkovic, S. J.; Kim, H.-S. *Science* **2006**, *311*, 535-538.

Chapter 5

Biomaterials Design: Human Tropoelastin

Introduction

Elastin is extremely important in a variety of tissues, *in vivo*, as the protein provides elasticity and resilience in organs such as the lung and skin as well as in the arteries of the cardiovascular system. However, due to the unique physical properties of native elastin, the structural protein is extremely insoluble above the inverse temperature transition characteristic of elastin and elastin-mimetic polypeptides[1]. Tropoelastin is the natural, soluble precursor of native elastin and is an interesting target for biomaterials applications[2]. Several human diseases are characterized by altered distribution of elastin and defective pathways related to elastin biosynthesis in the body (Table 1)[1]. Tropoelastin, as the natural, soluble precursor for native elastin, would be useful in treatments for elastin disorders by injection of recombinant tropoelastin in patients suffering from these disorders. Another application of tropoelastin as a biomaterial may include injection of soluble tropoelastin in damaged tissue as a treatment for injury or to promote wound healing. Soluble tropoelastin produced in bacteria at a high yield and at a low cost would provide large quantities of tropoelastin for biomaterials applications. Tropoelastin also has potential for use in anti-aging applications and in structural studies involving proline analogue incorporation as proline is abundant in native tropoelastin and plays a critical role in the structural stability of related elastin-mimetic polypeptides[3-5].

Although evidence suggests that only a single elastin gene is present in the mammalian genome[6], multiple tropoelastin isoforms in the human body exist due to alternative splicing of the 34 exons that comprise the gene[1,7-10]. 11 human tropoelastin splice variants have been identified, but tissue specificity has not been well characterized[1]. In addition, the functional roles of the various isoforms have not been

Disease	Features	Etiology
Atherosclerosis	Fragmentation of elastin in arteries, increased stiffness of arteries, increased lipid and calcium accumulation in elastin fibers	Uncertain, complex
Buschke-Ollendorff syndrome	Increased thick elastic fibers, decrease in microfibrils, skin lesions, bone dysplasia	Unknown, genetic
Cutis laxa (inherited)	X-linked and autosomally inherited forms, loose sagging skin, increased elastic fiber fragmentation, decreased lysyl oxidase activity	Uncertain, tropoelastin mRNA stability or copper transport defect proposed
Emphysema	Increased compliance of lung, loss of elastin in lung	Unbalanced protease/anti-protease activity suspected
Marfan syndrome	Widespread skeletal, ocular, and cardiovascular defects, loose skin, increased fragmentation of elastin	Mutations in fibrillin gene
Menkes syndrome	X-linked, brittle hair, tortuous blood vessels, elastic fiber fragmentation, neurological defects	Defect in copper-transport ATPase Mc-1 gene
Pseudoxanthoma elasticum	Inelastic skin, cardiovascular defects, fragmentation, clumping, and calcification of elastic fibers, increase in glycosaminoglycans	Unknown, elastin gene defect excluded
Supravalvular aortic stenosis (SVAS)	Narrowing of arteries, reduced elastin content, architecture of aorta disrupted	Mutations in elastin gene
Williams syndrome	SVAS, mental retardation, premature aging of skin, lax joints	Deletion of elastin gene allele and adjacent loci

Table 1. Diseases of elastin fibers[1]

elastin product and from elastin-mimetic polypeptides[9,11]. Two major types of domains are found in tropoelastin including hydrophobic domains rich in non-polar amino acids and hydrophilic domains rich in lysine residues (Table 2)[1]. The hydrophobic domains, believed to be important for the elastic properties of tropoelastin, are characterized by three to six peptide repeats with sequences such as GVGVP, GGVP, and GVGVAP. The hydrophilic domains, involved in cross-linking, consist of stretches of lysine residues separated by alanine residues such as in the sequence AAAKAAKAAA. In some tropoelastin isoforms, proline is present in place of a polyAla tract[11]. The hydrophilic and hydrophobic domains alternate in native tropoelastin. The hydrophobic domain sequences among species are evolutionarily divergent, while the hydrophilic, cross-linking domain sequences have been highly conserved over time[10].

In the human body, tropoelastin is synthesized with an N-terminal 26 amino acid secretion signal sequence that allows for secretion of tropoelastin from the cells as a ~72 kDa protein[11]. Deposition of tropoelastin in the extracellular matrix is a highly complex process that involves protein binding factors and subsequent fibrillogenesis[1]. Coacervation is one interesting property of soluble tropoelastin that is believed to be important for fibril formation. Upon increasing the temperature of a solution of soluble tropoelastin above 20 ºC, the solution becomes turbid due to aggregation of tropoelastin molecules. Aggregation is thermodynamically controlled, and is therefore fully reversible by cooling the solution[12]. Coacervation is believed to be caused by interaction between the hydrophobic domains, and plays an important role in fibrillogenesis[12-15].

Domain	Amino Acid Sequence
Hydrophobic domain (Exon 24 of human tropoelastin)	GLVPGVGVAPGVGVAPGVGVAPGVGLAPGVG VAPGVGVAPGVGVAPGIGPGGVA
Hydrophilic, cross-linking domain (Exon 19 of human tropoelastin)	GVVSPEAAAKAAAKAAKY

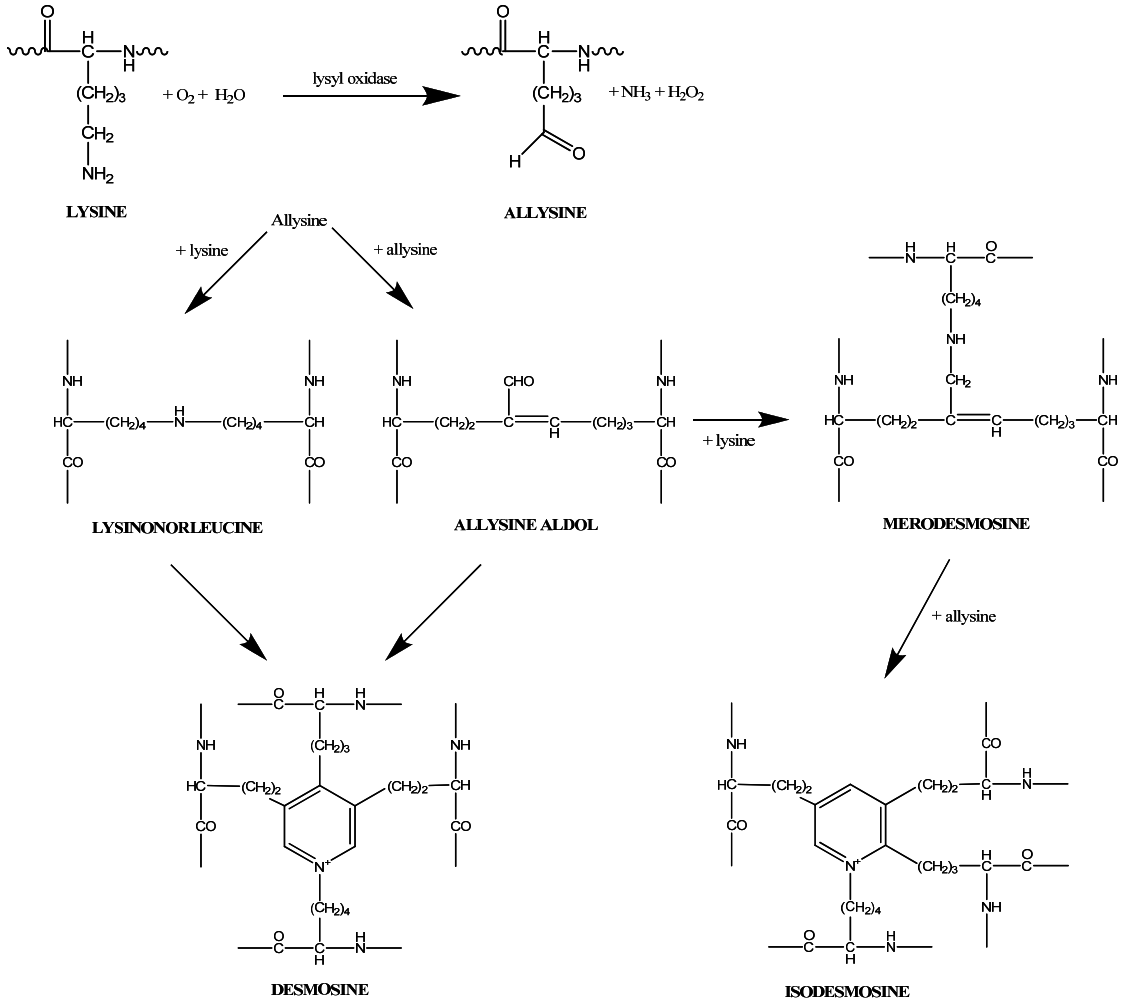
Table 2. The primary sequence of tropoelastin reveals alternating hydrophobic and hydrophilic domains. The hydrophobic domains, believed to be important for the elastic properties of tropoelastin, are characterized by three to six peptide repeats with sequences such as GVGVP, GGVP, and GVGVAP. The hydrophilic domains, involved in cross-linking, consist of stretches of lysine residues separated by alanine residues such as in the sequence AAAKAAKAAA[1].

After secretion, inter and intra-molecular cross-linking between lysine residues is carried out rapidly in the extracellular matrix by the copper-dependent enzyme, lysyl oxidase[16,17]. The enzyme targets most Lys residues in the tropoelastin sequence for cross-link formation. The enzyme-mediated cross-linking reaction involves oxidative deamination of a Lys residue to form allysine. Subsequent spontaneous condensation reactions follow to link allysine and nearby lysine residues and form cross-links such as allysine aldol, lysinonorleucine, merodesmosine, and unique tetrafunctional cross-links such as desmosine and isodesmosine (Scheme 1)[1,18]. Cross-linking of tropoelastin results in the formation of mature, native elastin. Lysyl oxidase is responsible for forming cross-links in native collagen as well[17]. Extraction and purification of tropoelastin from animal sources involves deprivation of copper from the diet of the animal leading to a loss in activity of lysyl oxidase and accumulation of tropoelastin in animal tissues[19,20]. Lysyl oxidase shows a preference for insoluble tropoelastin over soluble monomeric tropoelastin in solution giving evidence for the importance of coacervation in the biosynthesis of native elastin[21].

Biosynthesis and expression of human tropoelastin is an interesting goal for using tropoelastin in biomaterial applications and in structural cross-linking studies. The total synthesis and expression of human tropoelastin has been carried out using an *E. coli* host[22] and in yeast[23] by Anthony Weiss at the University of Sydney in Sydney, Australia. The group codon optimized the human gene for expression in bacteria as the gene (cDNA) contained many codons with a low frequency of usage in *E. coli*[22]. They constructed the gene using a series of oligonucleotides cloned in sequence into the pBluescript II SK^+ or pET3d vectors for expression. The ~65 kDa tropoelastin protein

was expressed and purified to a level of approximately 30 mg protein/liter culture in the pET3d vector. The isolation of human tropoelastin was followed by a series of studies addressing the coacervation characteristics of tropoelastin and the importance of the hydrophilic and hydrophobic domains on the structure of tropoelastin[24-31].

Amino acid analogue incorporation is one area of interest for biomaterials applications of tropoelastin. Extensive studies have demonstrated the effects of amino acid analogue incorporation on the structure of elastin-mimetic[4,5] and collagen-mimetic[32-34] polypeptides. Global dehydroproline incorporation, in tropoelastin in an *in vivo* study, resulted in an inhibition of tropoelastin secretion in embryonic chick cells[35]. Hydroxyproline present in tropoelastin, as a post-translational modification by the enzyme, prolyl hydroxylase, can result in a decrease in elastin production, possibly as a result of destabilization of the soluble structure of tropoelastin[36]. Our initial investigations were focused on high-level expression and purification of recombinant human tropoelastin for biomaterial applications or for use in structural studies involving analogue incorporation or cross-linking of tropoelastin.



Scheme 1. Structure and formation of lysine cross-links in elastin. After conversion of lysine to allysine by the enzyme, lysyl oxidase, spontaneous reaction of allysine with other lysine residues generates the cross-links depicted above[1].

Materials

All chemical reagents were purchased from either Fisher Scientific, Inc. (Pittsburgh, PA) or Sigma Chemical Co. (St. Louis, MO) unless otherwise noted. Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases, T4 DNA ligase, and T4 kinase were purchased from New England Biolabs, Inc. (Beverly, MA), shrimp alkaline phosphatase was obtained from Roche Applied Science (Indianapolis, IN), and Platinum *Pfx* DNA polymerase was obtained from Invitrogen Corp. (Carlsbad, CA). The *E. coli* CAG18515 and UQ27 strains were obtained from the *E. coli* Genetic Stock Center at Yale University. An auxotrophic phenotype was conferred to the UQ27 strain as described in Chapter 2 to create the $UQ27^{ts,\ ProA}$ strain. The *E. coli* strain, Top10F′ was obtained from Invitrogen Corp. (Carlsbad, CA).

Plasmids pQE-60 and pQE-80L were obtained from QIAGEN, Inc. (Valencia, CA). The plasmid, pBAD/HisA, was purchased from Invitrogen Corp. (Carlsbad, CA), and the pCDF-1b vector was purchased from Novagen, Inc. (Madison, WI). The plasmid pRAM2 was employed as a source of the gene encoding the **elastin-1** sequence. Double digestion of pRAM2 with *Nco* I and *Bam*H I afforded a duplex DNA cassette of approximately 1300 bp, which was inserted into the compatible *Nco* I/*Bgl* II sites of plasmid pQE-60 (QIAGEN, Inc., Valencia, CA). to generate plasmid pAG1. An *Eco*R I/*Hin*D III cassette derived from pAG1 was excised, isolated, and cloned into the compatible sites of plasmid pQE-80L (QIAGEN, Inc., Valencia, CA) to afford pAG2. Synthetic oligonucleotides were purchased from either Sigma-Genosys, Inc (The Woodlands, TX) or Integrated DNA Technologies (Coralville, IA) and were used as

received. TALON® metal affinity resin was purchased from Clontech, Inc. (Mountain View, CA). NMM medium was prepared as described previously[3], with the exception that proline was not added to the medium prior to cell culture. Similarly, SMM media was prepared according to the protocol by Studier[37].

Procedures for the manipulation of DNA, the transformation of competent cells, and the growth and induction of bacterial cultures were adapted from the published literature[38] or instructions supplied by manufacturers. Reagents for the manipulation of bacteria and DNA were sterilized by either autoclave or passage through a 0.2 μm filter. Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. Site-directed mutagenesis was performed using Stratagene's (La Jolla, CA) Quick-Change mutagenesis technique from gene-specific oligonucleotide primers. Automated DNA sequencing was performed at the Emory University Core DNA Sequencing Facility on a Perkin-Elmer ABI Prism model 377 DNA sequencer.

Protein electrophoresis was performed using 10-15 % gradient discontinuous pre-cast sodium dodecyl sulfate (SDS) polyacrylamide gels on a PhastSystem from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ). The Perfect Protein Marker was used as a protein standard for SDS-PAGE and western blot analysis and was purchased from Novagen, Inc. (Madison, WI). Silver staining, used for protein visualization, was carried out using a silver staining kit purchased from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ). Western blotting was carried out using the PhastSystem (GE Healthcare). The primary antibody used for binding his-tagged proteins was the His-tag monoclonal antibody from Clontech, Inc. (Mountain View, CA). The primary antibody for binding elastin or tropoelastin proteins was the mouse anti-elastin monoclonal

antibody from Chemicon International, Inc. (Millipore, Temecula, CA)[39]. The secondary antibody used was the goat anti-mouse secondary antibody, also from Clontech, Inc. Polyhistidine tagged proteins were visualized by chemiluminescent detection using the 1-step NBT/BCIP reagent mixture from Pierce Biotechnology (Thermo Scientific) (Rockford, IL).

Methods

Plasmid Construction and Cloning

The tropoelastin-containing plasmid DNA was isolated by transformation into the Top10F´strain and subsequent plasmid miniprep followed using the QIAGEN spin miniprep kit. The tropoelastin plasmid and the acceptor plasmid, pQE-60, were double-digested with *Nco* I/*Hind* III restriction endonucleases. The tropoelastin gene was then ligated into the acceptor plasmid in a reaction with DNA ligase. The sequence of the tropoelastin gene, in the new plasmid, pHC34, was confirmed by automated DNA sequencing analysis using primers designed to bind to the pQE-60 plasmid upstream and downstream of the multiple cloning site.

The new plasmid, pHC34 containing the tropoelastin gene, was then double-digested with *Eco*R I/*Hind* III, and the tropoelastin gene was isolated by agarose gel extraction. The digested tropoelastin gene was then ligated into the *Eco*R I/*Hind* III double-digested acceptor plasmid, pQE-80L. The plasmid was also isolated by gel extraction and dephosphorylated using a phosphatase enzyme prior to ligation. The sequence of the tropoelastin gene, in the new plasmid, pHC35, was confirmed by automated DNA sequencing analysis using primers designed to bind to the pQE-80L plasmid upstream and downstream of the multiple cloning site. The pQE-80L plasmid

provides an advantage over the use of the pQE-60 plasmid as the pQE-80L plasmid contains a copy of the overproducing repressor allele, $lacI_q$, to ensure tight control of the basal level of transcription prior to induction of tropoelastin protein expression with IPTG.

Tropoelastin was also cloned directly into the pQE-80L plasmid for fusion of a hexahistidine tag at the N-terminus of the tropoelastin gene to facilitate western blotting and metal-affinity column purification. Primers were designed to amplify the tropoelastin gene with a *Bam*H I restriction site at the 5′ terminus and a *Hind* III site at the 3′ terminus of the gene. The tropoelastin gene was amplified using PCR and the new primers, and the gene was subsequently double-digested with *Bam*H I/*Hind* III enzymes followed by agarose gel extraction. The digested tropoelastin gene was then ligated into the *Bam*H I/*Hind* III double-digested acceptor plasmid, pQE-80L. The plasmid was isolated by gel extraction and dephosphorylated using a phosphatase enzyme prior to ligation. The sequence of the tropoelastin gene, in the new plasmid, pHC46, was confirmed by automated DNA sequencing analysis using primers designed to bind to the pQE-80L plasmid upstream and downstream of the multiple cloning site.

The plasmid, pHC35 containing the tropoelastin gene, was then double-digested with *Eco*R I/*Hind* III, and the tropoelastin gene was isolated by agarose gel extraction. The digested tropoelastin gene was then ligated into the *Eco*R I/*Hind* III double-digested acceptor plasmid, pBAD/HisA. The plasmid was also isolated by gel extraction and dephosphorylated using a phosphatase enzyme prior to ligation. The sequence of the tropoelastin gene, in the new plasmid, pHC39, was confirmed by automated DNA sequencing analysis using primers designed to bind to the pBAD/HisA plasmid upstream

and downstream of the multiple cloning site. The pBAD/HisA derived plasmid, pHC39, contains an ampicillin resistance gene for selection and the tropoelastin gene under the control of the arabinose inducible promoter, araBAD. The arabinose inducible promoter allows for gradient-level protein expression with addition of varying concentrations of the arabinose inducer. The araBAD promoter was used to investigate the effects of variable levels of tropoelastin expression in the host compared to the highly active IPTG inducible promoter in the pQE-80L vector. The pBAD/HisA plasmid also incorporated a hexahistidine tag at the N-terminus of the tropoelastin gene to facilitate western blotting and metal-affinity column purification.

In order to increase expression levels of tropoelastin, the tropoelastin gene was cloned into the pCDF-1b plasmid under the control of the T7 promoter and the *lac* operator sequence. The plasmid, pHC34 containing the tropoelastin gene, was double-digested with *Nco* I/*Hind* III, and the tropoelastin gene was isolated by agarose gel extraction. The digested tropoelastin gene was then ligated into the *Nco* I/*Hind* III double-digested acceptor plasmid, pCDF-1b. The plasmid was also isolated by gel extraction and dephosphorylated using a phosphatase enzyme prior to ligation. The sequence of the tropoelastin gene, in the new plasmid, pHC45, was confirmed by automated DNA sequencing analysis. Cells expressing tropoelastin in the pCDF-1b background, also required co-expression of the T7 RNA polymerase harbored by the pMDT^{wt} plasmid (a gift from Dr. Shawn Desai at Emory University). The pMDT^{wt} plasmid has a spectinomycin resistance marker for plasmid maintenance.

Protein Expression

Small-scale expression in the UQ27$^{ts,\ ProA}$ strain*

pHC35 plasmid DNA (pQE-80L containing the **tropoelastin** gene) was transformed into the UQ27$^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL) and ampicillin (100 μg/mL). pHC35 plasmid DNA was also co-transformed with pHC15 plasmid DNA (containing the recombinant, wild-type *M. jannaschii* prolyl-tRNA synthetase/tRNA^{Pro} gene), or other appropriate plasmids, into the UQ27$^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). Five mL LB media, supplemented with appropriate antibiotics, was inoculated with single colonies of UQ27$^{ts,\ ProA}$ cells, harboring the appropriate plasmids, and grown overnight at 30 °C on a rotator. Two mL of the overnight cultures were transferred to 48 mL LB (supplemented with appropriate antibiotics) that had been added to 250 mL Erlenmeyer flasks for a total volume of 50 mL culture.

The 50 mL cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), protein expression was induced with IPTG. For induction, 50 uL, 1M IPTG were added to the flasks for a final concentration of 1 mM IPTG. The cultures were incubated at 30 °C for 3 h. OD_{600} readings were monitored for each of the cultures. 1 mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 μL ddH_2O and stored at –20 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 μL sample culture, 12.5 μL 2 x sample buffer, 1 μL DTT (dithiothreitol), and 6.5 μL ddH_2O for a total of 25 μL. Samples were placed at 95 ºC for 5 min.

*For expression in cell lines other than the UQ27 strain (CAG18515), the only modification from the outlined procedure is that cells were grown and incubated before and after induction at 37 ºC. For expression of plasmid pHC39 (tropoelastin in pBAD/HisA), protein expression was induced with varying concentrations of aqueous arabinose as indicated in the results section.

Large-scale expression

pHC35 plasmid DNA (pQE-80L containing the **tropoelastin** gene) was transformed into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL) and ampicillin (100 μg/mL). pHC35 plasmid DNA also was also co-transformed with pHC15 plasmid DNA (containing the recombinant, wild-type *M. jannaschii* prolyl-tRNA synthetase/$tRNA^{Pro}$ gene), or other appropriate plasmids, into the $UQ27^{ts,\ ProA}$ strain and plated on solid LB media supplemented with kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and ampicillin (100 μg/mL). Five mL LB media, supplemented with appropriate antibiotics were inoculated with single colonies of $UQ27^{ts,\ ProA}$ cells, harboring the appropriate plasmids, and grown overnight at 30 °C on a rotator. Five mL of the overnight cultures were transferred to 495 mL LB (supplemented with appropriate antibiotics) that had been added to 2,800 mL Erlenmeyer flasks for a total volume of 500 mL culture. Two flasks were used to express 1000 mL (1L) total volume of culture.

The large cultures were allowed to grow at 30 ºC to an OD_{600} ~0.7 over a period of 2-3 h. The OD_{600} was checked every hour. Once the cells reached log phase growth (OD_{600} ~0.7), protein expression was induced with IPTG. For induction, 500 μL, 1M IPTG were added to the flasks for a final concentration of 1 mM IPTG. The cultures

were incubated at 30 °C for 4 h. OD_{600} readings were monitored for each of the cultures. 1 mL aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged for 5 min at 6,300 rpm, the supernatant was discarded, and samples were resuspended in 50 µL ddH_2O and stored at –20 °C. After 3 h of induction, the two large cultures (1 L total volume) were spun down at 4,000 x g for 20 min. The cell pellets were resuspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and stored at –80 °C. Samples were prepared for SDS-PAGE and western blot analysis by mixing 5 µL sample culture, 12.5 µL 2 x sample buffer, 1 µL DTT (dithiothreitol), and 6.5 µL ddH_2O for a total of 25 µL. Samples were placed at 95 ºC for 5 min.

Purification of tropoelastin

A large-scale expression was carried out to obtain a large amount of the tropoelastin protein in the expression strain, $UQ27^{ts,\ ProA}$. Tropoelastin was expressed using the pHC35 expression plasmid. The procedure for purification of tropoelastin was modified from the procedure outlined by Martin *et al*[22]. The frozen cells were lysed by three freeze/thaw cycles. Lysozyme (1 mg/mL), protease inhibitor cocktail, benzonase (25 units/mL culture), and $MgCl_2$ (1 mM) were added to the lysate and the mixture was incubated with shaking (225 rpm) at 4 °C, overnight. The cell lysate was centrifuged at 10,000 x g for 30 min at 4 °C. The supernatant and pellet were separated and analyzed by SDS-PAGE to determine the location of the target protein. SDS-PAGE analysis indicated that the majority of the target protein was present in the soluble fraction with some residual protein in the pellet.

Two and a half volumes of 1-butanol were added to the aqueous supernatant from cell lysis to precipitate the tropoelastin protein. Aliquots of 1-butanol were added over a

2 h period in the cold room at 4 °C with constant stirring. A white milky precipitate was observed upon addition of 1-butanol. The suspension was centrifuged for 30 min at 11,000 x g, the alcohol fraction was extracted with a glass pipette. The remaining aqueous fraction, which contained tropoelastin, was added to a pre-massed, 500 mL round bottom flask and the solvent was removed under vacuum using a rotary evaporator. A white precipitate was observed on the inside of the flask. The flask was weighed to determine the mass of the resultant product and the white solid was resuspended in 20 mL 50 mM HEPES buffer at pH 8.0. SDS-PAGE analysis was conducted to determine the purity of the tropoelastin product.

Purification of tropoelastin was also carried out using a method that is useful for purification of elastin-mimetic proteins. A large, 1L, expression of tropoelastin (in pQE-80L) was carried out in the expression strain, $UQ27^{ts,\ ProA}$. The frozen cells were lysed by three freeze/thaw cycles. Lysozyme (1 mg/mL), protease inhibitor cocktail, benzonase (25 units/mL culture), and $MgCl_2$ (1 mM) were added to the lysate and the mixture was incubated with shaking (225 rpm) at 4 °C, overnight. The cell lysate was centrifuged at 10,000 x *g* for 30 min at 4 °C. The supernatant and pellet were separated and analyzed by SDS-PAGE to determine the location of the target protein. SDS-PAGE analysis indicated that the majority of the target protein was present in the soluble fraction with some residual protein in the pellet.

For the first hot/cold spin, the supernatant from the cell lysate was centrifuged at 10, 000 x *g* for 30 min at 4 °C. Aliquots of a 5 M NaCl solution were added to the supernatant containing the protein and the solution was incubated at 37 °C with shaking at 225 rpm for 1 h. A white precipitate appeared. The suspensions were then centrifuged

at 10, 000 x *g* for 30 min at 37 °C. The supernatant was collected for SDS-PAGE analysis and the pellet was resuspended in lysis buffer (pH 8.0) with protease inhibitor cocktail and incubated at 4 ºC with shaking for 12-16 h to re-dissolve the tropoelastin protein in the pellet.

The hot/cold spins were repeated two more times after the initial spins for a total of three hot/cold spins. Dialysis of the purified tropoelastin was carried out in sterile H_2O to remove residual salts and buffer. 10,000 MW tubing was used for the dialysis. The tropoelastin protein samples from dialysis were concentrated to a volume of 2 mL using Amicon Centriplus centrifugal filter devices (Millipore).

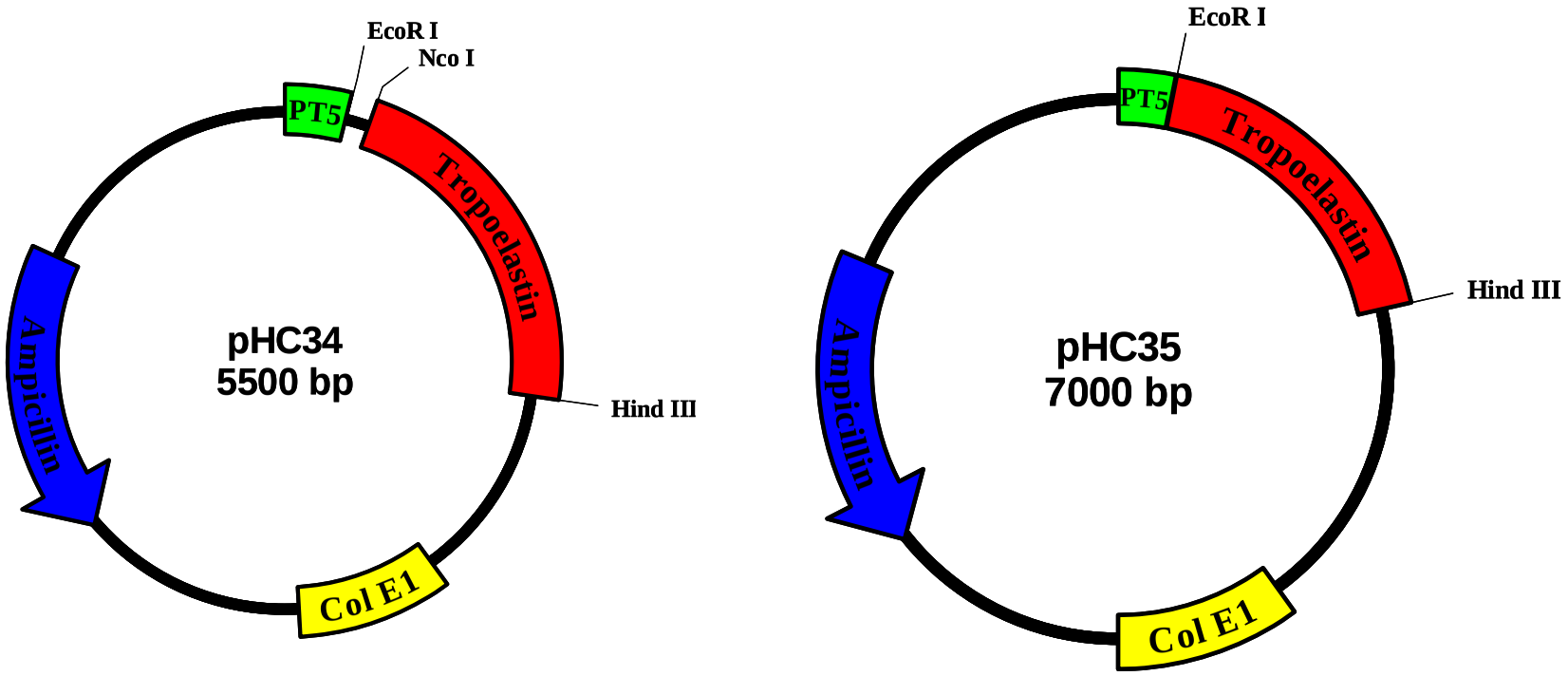


Figure 1. Plasmids constructed for the expression of recombinant human tropoelastin in a bacterial host. The pHC34 plasmid was derived from the pQE-60 vector, and the pHC35 plasmid was generated by sub-cloning of the tropoelastin gene cassette from the pHC34 plasmid using the *Eco*R I and *Hind* III restriction endonuclease cleavage sites in the multiple cloning regions of both plasmids.

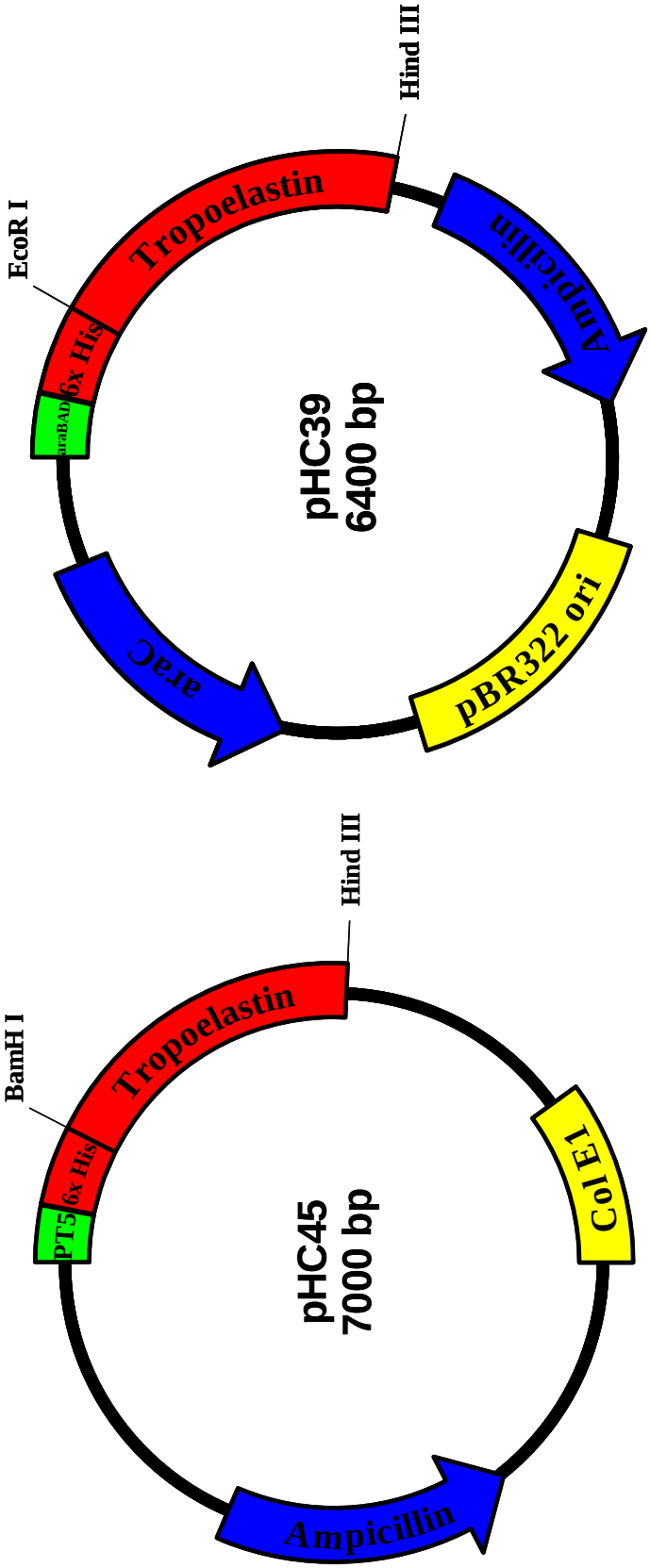


Figure 2. Plasmids constructed for the expression of recombinant human tropoelastin in a bacterial host. The pHC45 plasmid was derived from the pQE-80L vector, and the pHC39 plasmid was derived from the pBAD/HisA vector. Both plasmids incorporate a hexahistidine tag at the N-terminus of the tropoelastin gene to facilitate protein purification with a metal affinity column or western blotting analysis.

Results and Discussion

The human tropoelastin gene was designed and synthesized with codon optimization for high-level expression in an *E. coli* host. The complete tropoelastin gene was constructed by Blue Heron Technology, Inc. (Bothell, WA). Restriction endonuclease cleavage sites were incorporated at the 5′ and 3′ termini of the gene that was delivered in the pUC minus MCS plasmid supplied by the company. The human tropoelastin gene of interest is 2289 bp in length (Appendix 1) and was designed based on the known tropoelastin coding sequence from *H. sapiens*. The recombinant tropoelastin gene sequence is characterized by the hydrophilic and hydrophobic block domains characteristic of native tropoelastin.

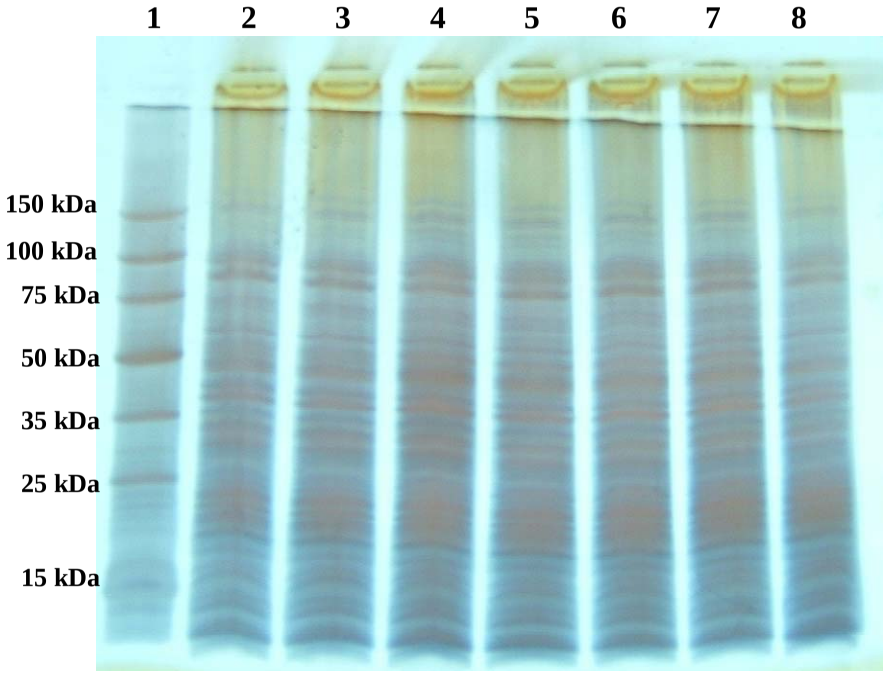
Expression of human tropoelastin was carried out in three different vector plasmid systems to determine the optimal conditions for high-level protein expression. Expression in the pCDF-1b vector resulted in the highest protein expression levels, while expression in the pQE80-L and pBAD vectors resulted in low protein expression yields. The higher product yields obtained with the pCDF-1b vector may have been due to the high-expression level achieved by using the T7 promoter. The Weiss group achieved high levels of tropoelastin expression in a pET vector also under the control of a T7 promoter[22]. However, we were interested in tighter control over expression of the tropoelastin gene for use in analogue incorporation studies.

Expression of tropoelastin was carried out in the pBAD vector system with arabinose induction in both LB and in minimal media to control glucose levels during expression. A series of arabinose concentrations were used to induce protein expression to determine if high tropoelastin levels in the cells were leading to toxicity and low

protein yields. No tropoelastin was detected by western blot or by SDS-PAGE analysis of expression samples from pBAD expressions (Fig. 3). The pQE-80L plasmid offered tighter control over expression levels and detectable levels of tropoelastin expression, so this plasmid was used to conduct expression and purification of tropoelastin (Fig. 4). Various concentrations of IPTG were used for induction but levels of tropoelastin expression remained constant at all final concentrations of IPTG (1 mM, 0.1 mM, 0.01 mM, 0.001 mM). The hexahistidine tagged tropoelastin gene in plasmid pHC45 (cloned directly into pQE-80L) was used for western blotting and could be purified using metal-affinity chromatography as an alternative to the modified procedure from the Weiss group[22] (Fig. 5). The anti-elastin antibody was fairly useful for western blotting with elastin-mimetic proteins, although several bands were visible on the blot possibly due to non-specific binding of the primary antibody to other proteins in the cell or to truncation products of elastin-mimetic protein translation. Use of the anti-elastin antibody was not effective for blotting tropoelastin samples possibly due to differences in the sequences of elastin-mimetics and tropoelastin proteins although the antibody was used previously for blotting of tropoelastin[39].

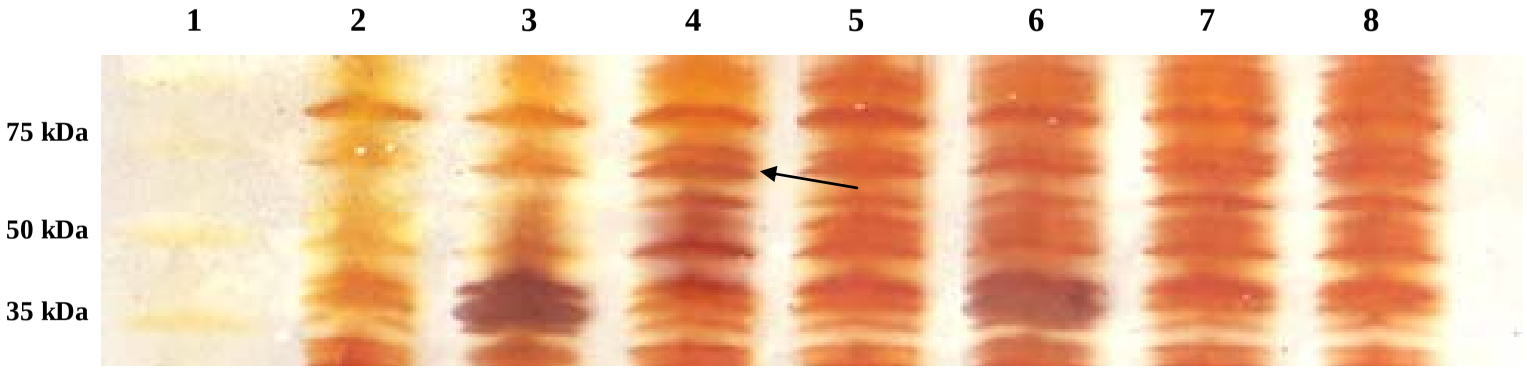
Purification of tropoelastin using the alcohol extraction method developed by Weiss[22] was fairly successful (Fig. 6). The protein yield was ~5 mg protein/L culture. Expression yield could be improved by using an alternative rich media such as TB or by expression using the pCDF-1b plasmid for high-level expression using the T7 promoter. Using an alternative expression host strain rather than the $UQ27^{ts,\ ProA}$ strain would likely result in higher yields. This strain was of interest due to its proline auxotrophic phenotype and its demonstrated efficiency in proline analogue incorporation. No

tropoelastin expression was detected from expression in the CAG18515 host. There were also small bands visible with SDS-PAGE analysis after purification of tropoelastin possibly due to cellular protein contaminants. Purification using the polyhistidine tagged tropoelastin in pQE-80L or optimization of the alcohol extraction method may result in a sample with fewer impurities.



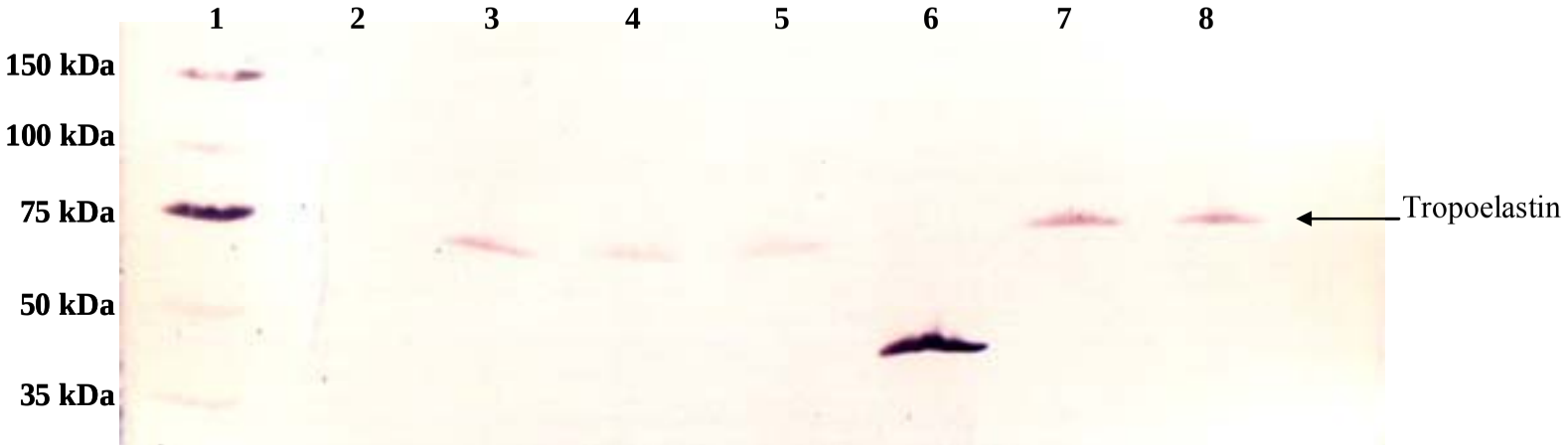
1. Perfect Protein Marker
2. Negative control, pBAD only
3. pHC39 (Tropoelastin in pBAD, 2% arabinose)
4. pHC39, 0.2 % arabinose
5. pHC39, 0.02 % arabinose
6. pHC39, 0.002 % arabinose
7. pHC39, 0.0002 % arabinose
8. pHC39, 0.00002 % arabinose

Figure 3. SDS-PAGE analysis with silver stain of cell extracts of tropoelastin expression samples. The pBAD vector did not allow for detectable levels of tropoelastin production with various final concentrations of arabinose used for induction. There are no tropoelastin bands visible above background levels of cellular expression on the gel. Tropoelastin should appear around 65-70 kDa.



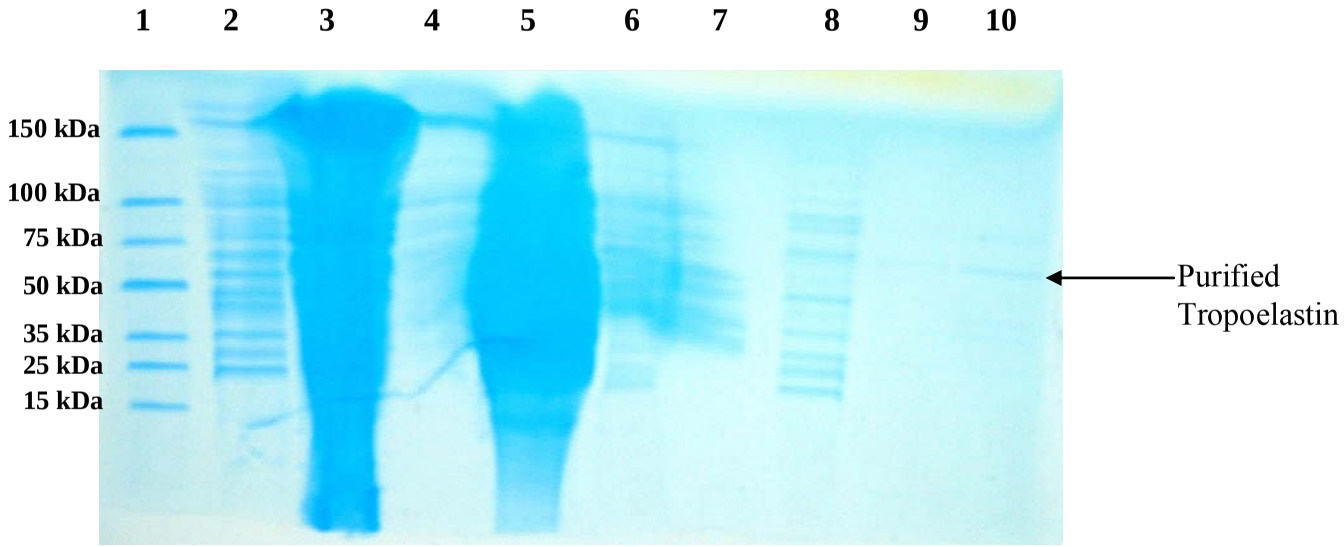
1. Perfect Protein Marker
2. Negative control, pQE-80L only, LB
3. pAG2 (**Elastin-1** in pQE-80L), LB
4. pHC35 (Tropoelastin in pQE-80L), LB
5. pQE-80L, Minimal media
6. pAG2 (**Elastin-1**), Minimal media
7. pHC35, Minimal media
8. pHC35, Minimal media

Figure 4. SDS-PAGE analysis with silver stain of cell extracts of tropoelastin expression samples. The pQE-80L vector produced detectable levels of tropoelastin production with IPTG induction in LB (Lane 4), but tropoelastin bands were not visible from expressions in minimal media indicating low expression levels. Expression of tropoelastin was carried out in rich LB media as well as in minimal media. Tropoelastin appears around 66 kDa, close to the predicted molecular mass. Expression levels of tropoelastin were low compared to **elastin-1** production possibly due to the lower growth temperature (30 °C) and the strain used for expression ($UQ27^{ts,\ ProA}$).



1. Perfect Protein Marker
2. Negative control, pQE-80L only
3. pHC45 (Tropoelastin in pQE-80L), LB
4. pHC45 (Tropoelastin in pQE-80L), LB
5. pHC45 (Tropoelastin in pQE-80L), LB
6. pHC8 (**Elastin-CCC**)
7. pHC45 (Tropoelastin in pQE-80L), Minimal media
8. pHC45 (Tropoelastin in pQE-80L), Minimal media

Figure 5. Western blot analysis of cell extracts of tropoelastin expression samples. The pQE-80L vector produced detectable levels of tropoelastin production with IPTG induction (Lanes 3, 4, 5, 7, 8). Tropoelastin appears around 66 kDa, close to the predicted molecular mass. Expression of tropoelastin was carried out in rich LB media as well as in minimal media. Expression levels of tropoelastin were low compared to **elastin-CCC** production possibly due to the lower growth temperature (30 °C) and the strain used for expression ($UQ27^{ts,\ ProA}$). Western blot analysis, a more sensitive method of detection, allowed for visualization of tropoelastin bands from expression in minimal media. The decahistidine tagged tropoelastin was useful for western blotting analysis.



1. Perfect Protein Marker
2. Supernatant 1 from cell lysis
3. Pellet 1 from cell lysis
4. Supernatant 2 from 1st Hot/Cold spin
5. Pellet 2 from 1st Hot/Cold spin
6. Tropoelastin from 1st Hot/Cold spin
7. Supernatant 3 from 2nd Hot/Cold spin
8. Pellet 3 from 2nd Hot/Cold spin
9. Pure tropoelastin
10. Pure tropoelastin (concentrated)

Figure 6. SDS-PAGE analysis of tropoelastin purification samples. The pQE-80L vector produced detectable levels of tropoelastin production with IPTG induction. Tropoelastin appears around 66 kDa, close to the predicted molecular mass. Expression of tropoelastin was carried out in rich LB media, but expression levels of tropoelastin were lower than expected possibly due to the lower growth temperature (30 °C) and the strain used for expression ($UQ27^{ts, ProA}$). The purified tropoelastin was visible on the gel (Lane 10), but the protein yield was low (~5 mg/L).

Conclusions

Recombinant human tropoelastin was expressed and purified from a bacterial protein expression host. The UQ27$^{ts,\ ProA}$ strain was sufficient for tropoelastin expression but expression in the host was low due to several factors. Growth at a lower temperature (30 °C) and expression with tightly controlled promoters were likely the main reasons for low protein expression and purification yields. Expression in the pCDF-1b vector resulted in the highest protein expression levels due to the high-level T7 promoter system. Purification using the polyhistidine tagged tropoelastin in pQE-80L or optimization of the alcohol extraction method would be useful as future goals for improving tropoelastin yields. Future studies with the purified tropoelastin may include cross-linking studies and incorporation of proline analogues to investigate the structure of tropoelastin. Chemically cross-linked tropoelastin may be useful in structural studies that investigate the in vivo folding of native elastin, or in biomaterials applications. Tropoelastin may be useful in a variety of biomaterials applications as elastin-mimetics have shown promise in biomedical device engineering and in a wide variety of other applications[40].

References

(1) Vrhovski, B.; Weiss, A. S. *European Journal of Biochemistry* **1998**, *258*, 1-18.

(2) Daamen, W. F.; Veerkamp, J. H.; van Hest, J. C. M.; van Kuppevelt, T. H. *Biomaterials* **2007**, *28*, 4378-4398.

(3) Kim, W. Y.; George, A.; Evans, M.; Conticello, V. P. *Chembiochem* **2004**, *5*, 928-936.

(4) Kim, W.; McMillan, R. A.; Snyder, J. P.; Conticello, V. P. *Journal of the American Chemical Society* **2005**, *127*, 18121-18132.

(5) Kim, W.; Hardcastle, K. I.; Conticello, V. P. *Angewandte Chemie-International Edition* **2006**, *45*, 8141-8145.

(6) Olliver, L.; Luvalle, P. A.; Davidson, J. M.; Rosenbloom, J.; Mathew, C. G.; Bester, A. J.; Boyd, C. D. *Collagen and Related Research* **1987**, *7*, 77-89.

(7) Fazio, M. J.; Mattei, M. G.; Passage, E.; Chu, M. L.; Black, D.; Solomon, E.; Davidson, J. M.; Uitto, J. *American Journal of Human Genetics* **1991**, *48*, 696-703.

(8) Bashir, M. M.; Indik, Z.; Yeh, H.; Ornsteingoldstein, N.; Rosenbloom, J. C.; Abrams, W.; Fazio, M.; Uitto, J.; Rosenbloom, J. *Journal of Biological Chemistry* **1989**, *264*, 8887-8891.

(9) Indik, Z.; Yeh, H.; Ornsteingoldstein, N.; Sheppard, P.; Anderson, N.; Rosenbloom, J. C.; Peltonen, L.; Rosenbloom, J. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 5680-5684.

(10) Boyd, C. D.; Christiano, A. M.; Pierce, R. A.; Stolle, C. A.; Deak, S. B. *Matrix* **1991**, *11*, 235-241.

(11) Indik, Z.; Yoon, K.; Morrow, S. D.; Cicila, G.; Rosenbloom, J.; Ornsteingoldstein, N. *Connective Tissue Research* **1987**, *16*, 197-211.

(12) Urry, D. W. *Journal of Protein Chemistry* **1988**, *7*, 1-34.

(13) Urry, D. W. *Scientific American* **1995**, *272*, 64-69.

(14) Urry, D. W. *Perspectives in Biology and Medicine* **1978**, *21*, 265-295.

(15) Wise, S. G.; Mithieux, S. M.; Raftery, M. J.; Weiss, A. S. *Journal of Structural Biology* **2005**, *149*, 273-281.

(16) Bressan, G. M.; Prockop, D. J. *Biochemistry* **1977**, *16*, 1406-1412.

(17) Kagan, H. M.; Sullivan, K. A. *Methods in Enzymology* **1982**, *82*, 637-650.

(18) Reiser, K.; McCormick, R. J.; Rucker, R. B. *Faseb Journal* **1992**, *6*, 2439-2449.

(19) Rucker, R. B. *Methods in Enzymology* **1982**, *82*, 650-657.

(20) Sandberg, L. B.; Wolt, T. B. *Methods in Enzymology* **1982**, *82*, 657-665.

(21) Soskel, N. T.; Wolt, T. B.; Sandberg, L. B. *Methods in Enzymology* **1987**, *144*, 196-214.

(22) Martin, S. L.; Vrhovski, B.; Weiss, A. S. *Gene* **1995**, *154*, 159-166.

(23) Toonkool, P.; Weiss, A. S. *Acta Biotechnologica* **2001**, *21*, 189-193.

(24) Vrhovski, B.; Jensen, S.; Weiss, A. S. *European Journal of Biochemistry* **1997**, *250*, 92-98.

(25) Wu, W. J.; Vrhovski, B.; Weiss, A. S. *Journal of Biological Chemistry* **1999**, *274*, 21719-21724.

(26) Jensen, S. A.; Vrhovski, B.; Weiss, A. S. *Journal of Biological Chemistry* **2000**, *275*, 28449-28454.

(27) Toonkool, P.; Jensen, S. A.; Maxwell, A. L.; Weiss, A. S. *Journal of Biological Chemistry* **2001**, *276*, 44575-44580.

(28) Toonkool, P.; Regan, D. G.; Kuchel, P. W.; Morris, M. B.; Weiss, A. S. *Journal of Biological Chemistry* **2001**, *276*, 28042-28050.

(29) Clarke, A. W.; Arnspang, E. C.; Mithieux, S. M.; Korkmaz, E.; Braet, F.; Weiss, A. S. *Biochemistry* **2006**, *45*, 9989-9996.

(30) Keane, F. M.; Clarke, A. W.; Foster, T. J.; Weiss, A. S. *Biochemistry* **2007**, *46*, 7226-7232.

(31) Muiznieks, L. D.; Weiss, A. S. *Biochemistry* **2007**, *46*, 8196-8205.

(32) Shoulders, M. D.; Raines, R. T. *Biopolymers* **2007**, *88*, 593-593.

(33) Kotch, F. W.; Guzei, I. A.; Raines, R. T. *Journal of the American Chemical Society* **2008**, *130*, 2952-2953.

(34) Shoulders, M. D.; Guzei, I. A.; Raines, R. T. *Biopolymers* **2008**, *89*, 443-454.

(35) Schein, J.; Frankel, L.; Rosenbloom, J. *Archives of Biochemistry and Biophysics* **1977**, *183*, 416-420.

(36) Urry, D. W.; Sugano, H.; Prasad, K. U.; Long, M. M.; Bhatnagar, R. S. *Biochemical and Biophysical Research Communications* **1979**, *90*, 194-198.

(37) Studier, F. W. *Protein Expression and Purification* **2005**, *41*, 207-234.

(38) Sambrook, J.; Russell, D. W. *Molecular Cloning: A laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

(39) Sato, F.; Wachi, H.; Starcher, B. C.; Murata, H.; Amano, S.; Tajima, S.; Seyama, Y. *Clinical Biochemistry* **2006**, *39*, 746-753.

(40) van Hest, J. C. M.; Tirrell, D. A. *Chemical Communications* **2001**, 1897-1904.

Appendix 1

Sequences of Interest

DNA sequence of the MjProRS enzyme

5′-ATGTTGGAATTTTCAGAATGGTATTCAGATATATTAGAAAAAGCTGAA

ATTTATGATGTTAGGTATCCAATAAAAGGTTGTGGAGTTTATTTACCTTACGG

ATTTAAAATAAGAAGATACACATTCGAAATAATAAGAAATTTATTAGATGAG

AGTGGGCATGATGAGGCATTATTCCCAATGCTGATTCCAGAGGATTTATTAGC

TAAGGAGGCAGAGCATATAAAAGGATTTGAGGATGAGGTTTATTGGGTAACT

CATGGAGGAAAAACACAGTTAGATGTTAAATTAGCTTTAAGACCTACTTCAG

AAACACCAATATACTATATGATGAAACTTTGGGTTAAGGTTCATACTGATTTG

CCAATAAAAATCTATCAGATAGTTAATACATTTAGGTATGAAACAAAGCACA

CAAGACCTTTAATTAGGTTAAGAGAGATAATGACATTTAAAGAGGCCCACAC

TGCCCATTCAACAAAGGAAGAGGCTGAAAACCAAGTAAAAGAAGCTATATCT

ATCTACAAAAAATTCTTTGATACTTTGGGTATTCCTTATTTAATATCCAAAAG

ACCAGAATGGGACAAATTCCCTGGGGCAGAATACACAATGGCTTTTGACACT

ATATTCCCAGATGGAAGAACTATGCAGATAGCTACAGTCCATAACTTAGGGC

AGAACTTCTCAAAGACATTTGAAATTATATTTGAAACACCAACTGGAGATAA

AGATTATGCTTATCAAACATGCTACGGAATCTCAGATAGGGTTATAGCTTCAA

TTATAGCAATACATGGGGATGAGAAAGGTTTAATTCTGCCTCCAATAGTTGC

ACCAATACAGGTAGTTATAGTTCCATTAATTTTCAAAGGAAAGGAAGATATT

GTTATGGAGAAGGCAAAAGAGATTTATGAGAAATTAAAAGGTAAATTTAGA

GTCCATATAGATGATAGGGACATAAGACCTGGAAGGAAGTTTAACGATTGGG

AGATAAAAGGCGTTCCATTGAGGATTGAAGTAGGTCCAAAAGATATTGAGAA

TAAAAAGATAACCTTATTTAGAAGAGATACAATGGAGAAATTCCAGGTGGAT

GAAACCCAGTTAATGGAGGTTGTAGAAAAAACTTTAAATAATATTATGGAAA

ACATTAAGAATAGAGCATGGGAAAAATTCGAAAACTTTATAACCATCCTTGA

AGATATAAATCCTGATGAAATTAAAAATATACTATCTGAAAAGAGGGGGGTA

ATTTTAGTCCCATTTAAGGAAGAGATATACAACGAAGAACTTGAAGAGAAAG

TAGAGGCAACTATTTTAGGGGAGACAGAATATAAAGGTAATAAATATATAGC

AATAGCTAAAACCTACTAA-3′

Amino acid sequence of the MjProRS enzyme, N-terminus to C-terminus

MLEFSEWYSDILEKAEIYDVRYPIKGCGVYLPYGFKIRRYTFEIIRNLLDESGHDE

ALFPMLIPEDLLAKEAEHIKGFEDEVYWVTHGGKTQLDVKLALRPTSETPIYYM

MKLWVKVHTDLPIKIYQIVNTFRYETKHTRPLIRLREIMTFKEAHTAHSTKEEAE

NQVKEAISIYKKFFDTLGIPYLISKRPEWDKFPGAEYTMAFDTIFPDGRTMQIATV

HNLGQNFSKTFEIIFETPTGDKDYAYQTCYGISDRVIASIIAIHGDEKGLILPPIVAPI

QVVIVPLIFKGKEDIVMEKAKEIYEKLKGKFRVHIDDRDIRPGRKFNDWEIKGVPL

RIEVGPKDIENKKITLFRRDTMEKFQVDETQLMEVVEKTLNNIMENIKNRAWEKF

ENFITILEDINPDEIKNILSEKRGVILVPFKEEIYNEELEEKVEATILGETEYKGNKYI

AIAKTY

ClustalW analysis of synthetase enzymes indicating homology at conserved residues

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
EcProRS         -MRTSQYLLS TLKETPADAE VISHQLMLRA GMIRKLASGL YTWLPTGVRV
MjProRS         ---------- -MEFSEWYSD ILEKAEIY-- -DVRYPIKGC GVYLPYGFKI
TtProRS         MAKEKGLTPQ SQDFSEWYLE VIQKAELA-- -DYG-PVRGT IVVRPYGYAI
ClustalW                     . :    : ::.:  :           *   .  * *  :


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
EcProRS         LKKVENIVRE EMNNAGAIEV SMPVVQPADL WQ-ESGRWEQ YGPELLRFVD
MjProRS         RRYTFEIIRN LLDESGHDEA LFPMLIPEDL LAKEAEHIKG FEDEVYWVTH
TtProRS         WENIQQVLDR MFKETGHQNA YFPLFIPMSF LRKEAEHVEG FSPELAVVTH
ClustalW         .   ::: .  :.::*  :.  :*:. * .:    *: : :  :  *:  ...
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
EcProRS         RG----ERPF VLGPTHEEVI TDLIRNELSS YKQLPLNFYQ IQTKFRDEVR
MjProRS         GGKTQLDVKL ALRPTSETPI YYMMKLWVKV HTDLPIKIYQ IVNTFRYETK
TtProRS         AGGEELEEPL AVRPTSETVI GYMWSKWIRS WRDLPQLLNQ WGNVVRWEMR
ClustalW         *    :  : .: ** *  *   :    :     :**  : *   . .* * :

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
EcProRS         PRFGVMRSRE FL-MKDAYSF HTSQESLQET YDAMYAAYSK IFSRMGLDFR
MjProRS         HTRPLIRLRE IMTFKEAHTA HSTKEEAENQ VKEAISIYKK FFDTLG-IPY
TtProRS         -TRPFLRTSE FL-WQEGHTA HATREEAEEE VRRMLSIYAR LAREYAAIPV
ClustalW            .:*  * ::  ::.::  *:::*. ::       : * : :    .

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
EcProRS         AVQADTGSIG GQASHEFQVL AQSGEDDVVF SDTSDYAANI ELAEAIAPKE
MjProRS         LISKRP---- --EWDKFPGA EYTMAFDTIF PDG------- ----------
TtProRS         IEGLKT---- --EKEKFAGA VYTTTIEALM KDG------- ----------
ClustalW             .         .:*      :   :.::  *

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
EcProRS         PRAAATQEMT LVDTPNAKTI AELVEQFNLP IEKTVKTLLV KAVEGSSFPQ
MjProRS         -RTMQIATVH NLGQNFSKTF EIIFETPTGD KDYAYQTCYG ISDR---VIA
TtProRS         -KALQAGTSH YLGENFARAF DIKFQDRDLQ VKYVHTTSWG LSWR---FIG
ClustalW         ::         :.   ::::    .:       . .  *     : .   .

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
EcProRS         VALLVRGDHE LNEVKAEKLP QVASPLTFAT EEEIRAVVKA GPGSLGPVNM
MjProRS         SIIAIHGDEK GLILPPIVAP IQVVIVPLIF KGKEDIVMEK AKEIY-----
TtProRS         AIIMTHGDDR GLVLPPRLAP IQVVIVPIYK DESRERVLEA AQGLR-----
ClustalW          :  :**..    : .   *   .  :.:   . .   *::  .

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       360        370        380        390        400
EcProRS         PIPVVIDRTV AAMSDFAAGA NIDGKHYFGI NWDRDVATPE VADIRNVVAG
MjProRS         ---------- EKLKGK-FRV HID------- --DRD-IRPG RKFNDWEIKG
TtProRS         ---------- QALLAQGLRV HLD------- --DRDQHTPG YKFHEWELKG
ClustalW                  :      . ::*          ***   *         : *

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       410        420        430        440        450
EcProRS         DPSPDGQGRL LIKRGIEVGH IFQLGTKYSE ALKASVQGED GRNQILTMGC
MjProRS         VPLR------ ---------- -IEVGPKDIE NKKITLFRRD ------TMEK
TtProRS         VPFR------ ---------- -VELGPKDLE GGQAVLASRL GGKETLPLAA
ClustalW         *                    .::*.*  *   :  :  .        .:
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       460        470        480        490        500
EcProRS         YGIGVTRVVA AAIEQNYDER GIVWPDAIAP FQVAILPMNM HKSFRVQELA
MjProRS         FQVDETQLME VVEKTLNN-- ---------- ----IMENIK NRAWEKFENF
TtProRS         LPEALPGKLD AFHEELYR-- ---------- ----RALAFR EDHTRKVDTY
ClustalW             .  :  .  :                              .   .  :


                ....|....| ....|....| ....|....| ....|....| ....|....|
                       510        520        530        540        550
EcProRS         EKLYSELRAQ GIEVLLDDRK ERPGVMFADM ELIGIPHTIV LGDRNLDNDD
MjProRS         ITILEDINPD EIKNILSEKR GVILVPFKEE IYN-----EE LEEKVEATIL
TtProRS         EAFKEAVQEG FALAFHCGDK ACERLIQEET TATTRCVPFE AEPEEGFCVR
ClustalW          : . :.       :    :     :   :                .


                ....|....| ....|....| ....|....
                       560        570
EcProRS         IEYKYRRNGE KQLIKTGDIV EYLVKQIKG
MjProRS         GETEYKGNKY IAIAKTY--- ---------
TtProRS         CGRPSAYGKR VVFAKAY--- ---------
ClustalW               .     : *:
```

ClustalW alignment of the amino acid sequences of GFPuv, superfolder GFP, and DsRed

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
GFPuv           MSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT
SuperGFP        MSKGEELFTG VVPILVELDG DVNGHKFSVR GEGEGDATNG KLTLKFICTT
DsRedExp        MASSEDVIKE FMRFKVRMEG SVNGHEFEIE GEGEGRPYEG TQTAKLKVTK
ClustalW        *:..*:::.  .: : *.::* .****:*.:  ***** .  * . * *:  *.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
GFPuv           G-KLPVPWPT LVTTFSYGVQ CFSRYPDHMK RHDFFKSAMP EGYVQERTIS
SuperGFP        G-KLPVPWPT LVTTLTYGVQ CFSRYPDHMK RHDFFKSAMP EGYVQERTIS
DsRedExp        GGPLPFAWDI LSPQFQYGSK VYVKHPADIP --DYKKLSFP EGFKWERVMN
ClustalW        *  **..*   * . : ** :  : ::* .:    *: * ::* **:  **.:.


                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
GFPuv           FKDDGNYKTR AEVKFEGDTL VNRIELKGID FKEDGNILGH KLEYNYNSHN
SuperGFP        FKDDGTYKTR AEVKFEGDTL VNRIELKGID FKEDGNILGH KLEYNFNSHN
DsRedExp        FEDGGVVTVT QDSSLQDGSF IYKVKFIGVN FPSDGPVMQK KTMGWEASTE
ClustalW        *:*.*  ..   : .::..:: : :::: *:: * .** :: : *      * :
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
GFPuv           VYITADKQKN GIKANFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH
SuperGFP        VYITADKQKN GIKANFKIRH NVEDGSVQLA DHYQQNTPIG DGPVLLPDNH
DsRedExp        RLYPRDGVLK GEIHKALKLK DGGHYLVEFK SIYMAKKPVQ LPGYYYVDSK
ClustalW           . *   : *   :    : :  .  *::  . *  :.*:         *.:


                ....|....| ....|....| ....|....| ....|....| ....|..
                       210        220        230        240
GFPuv           YLSTQSALSK DPNEKRDHMV LLEFVTAAGI THGMDELYK- -------
SuperGFP        YLSTQSVLSK DPNEKRDHMV LLEFVTAAGI THGMDELYKG SHHHHHH
DsRedExp        LDITS----- ---HNEDYTI VEQYERAEGR HHLFL----- -------
ClustalW           *.         .:.*: : : ::  * *   * :
```

Appendix 2

Laboratory Recipe for Making Modified Studier Minimal Media

Studier Minimal Media (SMM), Pro+ (1 Liter)

- 25 mM KH_2PO_4 25 mL 1 M stock solution (monobasic)
- 25 mM Na_2HPO_4 25 mL 1 M stock solution dibasic
- 50 mM NH_4Cl 50 mL 1 M stock solution
- 2 mM $MgSO_4$ 2 mL 1 M stock solution
- 5 mM Na_2SO_4 5 mL 1 M stock solution
- 20 mM glucose 20 mL 1 M stock solution
- 1 µg/mL trace metals (Zn^{2+}, Mn^{2+},Cu^{2+}, MoO_4^{2-}); 100 µL 100 mg/L stock solution
- 50 µg/mL all amino acids; 100 mL of 10 x stock solution, 500 mg each amino acid in 1 L PBS solution (pH=10 with NaOH pellets). Filter sterilized, store at 4 ºC with foil.
- 100 mL of a 500 mg proline in 1 L PBS solution
- Add 672.9 mL ddH_2O to 1 L
- Mix above and filter sterilize. Store at 4 ºC with foil.

Studier Minimal Media (SMM), Pro- (1 Liter)

- 25 mM KH_2PO_4 25 mL 1 M stock solution (monobasic)
- 25 mM Na_2HPO_4 25 mL 1 M stock solution dibasic
- 50 mM NH_4Cl 50 mL 1 M stock solution
- 2 mM $MgSO_4$ 2 mL 1 M stock solution
- 5 mM Na_2SO_4 5 mL 1 M stock solution
- 20 mM glucose 20 mL 1 M stock solution
- 1 µg/mL trace metals (Zn^{2+}, Mn^{2+},Cu^{2+}, MoO_4^{2-}); 100 µL 100 mg/L stock solution
- 50 µg/mL all amino acids; 100 mL of 10 x stock solution
- Add 772.9 mL ddH_2O to 1 L
- Mix above and filter sterilize. Store at 4 ºC with foil.