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.

_____

Melissa A. Patterson

Synthesis of protein-based polymers with the potential to form physically and covalently cross-linked networks

By

Melissa A. Patterson
Doctor of Philosophy

Department of Chemistry

_____

Vincent P. Conticello, Ph.D.
Adviser

_____

Elliot L. Chaikof, MD, Ph.D.
Committee Member

_____

Dale E. Edmondson, Ph.D.
Committee Member

Accepted:

_____

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

_____

Date

Synthesis of protein-based polymers with the potential to form physically and covalently cross-linked networks

By

Melissa A. Patterson
B.S., Oglethorpe University, 2004

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

2011

Abstract

Synthesis of protein-based polymers with the potential to form physically and covalently cross-linked networks

By Melissa Ann Patterson

Materials science is targeting the construction of 'smart' biomaterials where stimulus by physiologically relevant cues induces an event at the molecular level, which acts cooperatively to bring about a macromolecular response. The first part of the work described herein has focused on development of a cloning strategy for the biosynthesis of high molecular weight, multi-domain polymers, derived from the pentapeptide (VPGVG) repeat motif characterizing native elastin. An attractive strategy for protein-based hydrogel development involves genetically fusing dissimilar polypeptide blocks with unique temperature-dependent hydrophobic assembly behavior and mechanical properties. Development of a biosynthetic strategy that is modular and seamless is reported as well as the temperature-dependent phase behavior of amphiphilic block copolymers as determined by DSC.

The remainder of the dissertation describes work toward developing a series of *E. coli* expression strains that are competent for multi-site incorporation of unnatural amino acids. UAAs were selected because their incorporation would provide novel strategies for further derivatization or cross-linking of the elastin-mimetic polypeptides under physiologically relevant conditions. In Chapters 3 and 4, residue-specific incorporation strategies are described for incorporation of Hag and DOPA under conditions that are compatible with high levels of analogue replacement and relatively high yield. The fidelity of analogue substitution was examined by amino acid analysis, mass spectrometry, and NMR studies for qualitative and quantitative measurements of replacement of the natural amino acid. The temperature-dependent

assembly behavior of the analogue-containing and wild type elastin-mimetic polymers was investigated by DSC and temperature-dependent turbidity measurements. In addition, the metal-ligand binding and cross-linking behavior of the elastin-mimetic polymer containing DOPA was studied. Finally, Chapter 5 details the use of a site-specific incorporation method for multi-site incorporation of the benzophenone analogue of tyrosine, Bpa. Using an orthogonal suppressor tRNA/synthetase pair, the unnatural amino acid is incorporated into the polypeptide chain in response to the amber codon UAG. A mutant host strain is employed, which is characterized by a ribosomal mutation that alters release factor activity. The efficacy of the incorporation strategy is judged by protein yield, FACS analysis, and mass spectrometry.

Synthesis of protein-based polymers with the potential to form physically and covalently cross-linked networks

By

Melissa A. Patterson
B.S., Oglethorpe University, 2004

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

2011

Acknowledgements

I would like to thank my advisor, Dr. Vince Conticello, for his guidance, support, and encouragement from my first research experience in the lab as an undergraduate student through my years as a graduate student. I look forward to using his advice and the skills I've learned from him along my continued path in research. I would also like to extend my gratitude to my committee members, Dr. Dale Edmondson and Dr. Elliot Chaikof. I have always appreciated their thoughtful questions, enthusiasm for research, and taking time to help me improve my presentation and writing skills. Within the Emory Chemistry Department, I would like to thank the following individuals and research centers for use of their assistance and instrument use: Dr. Fred Strobel, from the Mass Spectrometry Center, Dr. Shaoxiong Wu and Dr. Bing Wang, from the NMR Research Center, the directors of the Robert P. Apkarian Integrated Electron Microscopy Core, and Dr. David Lynn. In addition, I would like to thank Ann Dasher for monitoring my progress in graduate school, Steve Krebs and Patti Barnett of the stockroom,for helping me to get materials for my experiments and providing friendly conversation & support, and Tim Stephens and Patrick Morris, for keeping our equipment in order. From the Emory community, I would also like to acknowledge Dr. Pat Marstellar and Jordan Rose of the Center for Science Education. They have both been fantastic mentors who have given advice on professional development, choosing a career path, reminding me to take stock of accomplishments, and engaging others in science.

During my time in the Conticello lab, I have been lucky to work with many talented, interesting, kind, and supportive people. Many thanks to Jaimie Anderson, Yuri Zimenkov, Wookhyun Kim, Jet Qi, Holly Carpenter, Sonha Payne, Steven Dublin, John Shugart, Yang Feng, Yunyun Pei, Ye Tian, Weilin Peng, I-Lin Wu, Chris Xu, Paolo Anzini, and Tao Jian. I

would like to make special mention of Dr. Sonha Payne whose recommendation allowed me to begin working in the lab and whose training instilled good habits at the bench and in the lab.

Thanks to all of my friends who have been avid cheerleaders, great listeners, and provided me relief from the occasional stresses of graduate school life. To Emma, Jenna, Cheryl, Dana, and Amy (best friends I could ask for); and Lorna, Erin & little Olive, Skip, Stacey & Jon, Mariana & Bobby, the Isenberg family (friends that feel like family); and many more fellow graduate students, roller girls, and cyclists (trackies, roadie, fixies, and tandemers!) I would also like to give special thanks to Billy for all the experiences we've enjoyed cycling, for his musical talent which fills the apartment with good tunes, a best-friend to share adventures in Atlanta over the years, and support at the end of the day. Special thanks goes to my family, my maternal grandparents Bonnie & Jim, my paternal grandparents Clarence & Frances (Papa Pat & Nana), my parents and my extended family. I am grateful to Papa Pat whose career as a chemist for ALCOA was an inspiration. And I am very grateful to Nana for her generosity of time, love, grace & good style! Although they are not here to celebrate the completion of this work, I was lucky to have their support during my first years as a graduate student and their pride & encouragement has remained with me. A special thanks to my parents, Tim and Sharon for all the experiences that they provided for me over the years that helped me to get here today. I appreciate all the time Mom took to educate me at home after school at an early age and to share her love of art and literature. To my Dad, who instilled in me a great sense of adventure and exploration through many miles on the bicycle and on the trail, I am grateful. I am always grateful to have their love and support.

Table of Contents

Synthesis of protein-based polymers with the potential to form physically and covalently cross-linked networks

List of Figures

Chapter 1.

Chapter 2.

Chapter 3.

List of Tables

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

List of Schemes

Chapter 1

Introduction

Adapted in part from Payne, S. C.; Patterson, M. A.; Conticello, V. P. In *Protein Engineering Handbook*, Lutz, S., Bornscheuer, U. T., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2009; Vol. 1, p 915-936.

Protein-based biomaterials

Over the past few decades, the *de novo* synthesis of protein-based materials has emerged as a burgeoning area of polymer research[1]. Genetic engineering of protein-based polymers is motivated by two distinct aims. The synthetic polymers can serve as models for (1) the study of architecturally uniform macromolecules and (2) the investigation of native protein-based materials as well as the generation of synthetic mimics. From a fundamental polymer science perspective, protein-based materials allow for an understanding of the effects of polymer architecture (chain length, sequence, stereochemistry, and side chain functionality) on macromolecular properties. Furthermore, protein biosynthesis is an attractive alternative to conventional synthetic methods (i.e. solid-phase synthesis), where uniformity of polymer microstructure is not ensured and synthetic polymer size is limited by decreasing yield with each successive coupling step. Genetic engineering allows for the synthesis of non-natural polypeptide sequences, which can serve as model uniform polymers. In addition, the observed control of polypeptide primary structure via protein biosynthesis may permit the formation of defined higher-order structures through the progression of protein structural hierarchy. Specifically, the design and selection of protein structure at the primary level may specify secondary and super-secondary elements, the interactions between them, and the self-assembly of polymers into supramolecular aggregates, such as structurally defined fibrils[2], thermoresponsive nanoparticles[1], and nanostructured hydrogels[3] (Figure 1[4]).

Research in protein-based materials also aims to understand the chemical, biological, and mechanical properties that characterize the biological function of native protein polymers. Native protein-based materials constitute a library of structurally diverse variants, which typically

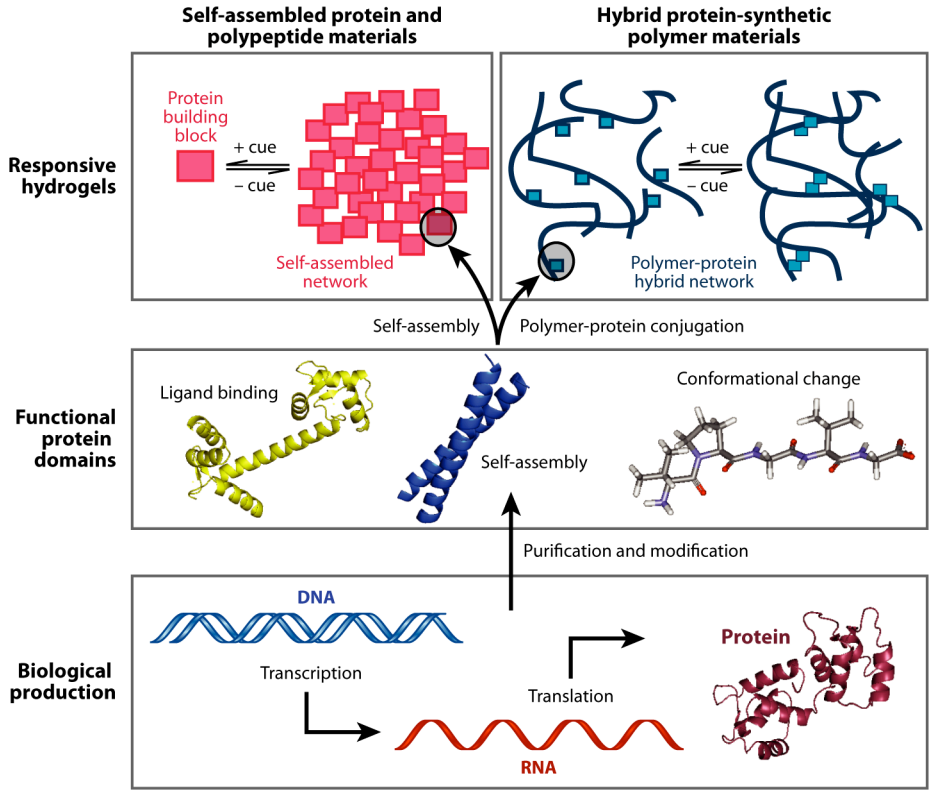


Figure 1. Protein engineering of responsive protein-based materials with controlled macromolecular architectures. Control of the primary amino acid sequence at the DNA level leads to control of the three-dimensional structure and function of the responsive protein domain or peptide motif. Fine-tuning of the structure and consequent function of the protein domain can be achieved at the DNA level, as mutation to the protein encoding sequence leads to specific changes in the primary amino acid sequence, and the secondary, tertiary, and quaternary structures. Three responsive peptide domains are illustrated: calmodulin in extended dumbbell conformation is an example of a ligand binding protein; a two-stranded leucine zipper domain is an example self-assembly domain; and an elastin-like peptide is an example of a conformational change domain.

display a low complexity sequence consisting of tandem repeats of a consensus oligopeptide motif[5]. Many of these proteins hold a position of particular importance as invaluable structural elements within biological systems (Figure 2). The primary structure of these native materials presumably determines their unique structural and functional properties. Furthermore, the sequence specificity strongly influences the mechanism of self-assembly of the polymer chains into the supramolecular architectures that characterize the materials properties and performance.

Protein-based polymer research has thereby been directed toward understanding the sequence-structure-function relationship that defines the biological function of native materials. Such insight would provide a rationale for design of artificial analogues with the potential to emulate or expand upon the functional utility of their natural counterparts. To date, recombinant protein expression has permitted the synthesis of a wide variety of synthetic protein-based materials derived from the canonical repeat sequences of native fibrous proteins such as elastin, collagen, keratin, and silk[6-8]. Such protein-based polymers are attractive targets for biomedical applications. Specifically, the ability to tailor the biological, chemical, and mechanical parameters of the polymer network through selection of primary sequence imparts an attractive technological advantage over conventional polymeric materials[9-12]. Furthermore, these protein-based polymers would be used for the generation of 'smart' biomaterials, where changes at the molecular level work cooperatively to bring about a well-defined, macroscopic response. This study describes the genetic engineering of protein-based polymers with applications as responsive biomaterials via the following two novel and complementary strategies: (1) a biosynthesis technique that is both *modular* and *convergent* permitting the generation of high molecular weight amphiphilic block copolymers and (2) the *high specificity, multi-site* incorporation of unnatural amino acids with useful side chain functionality during *in vivo* protein

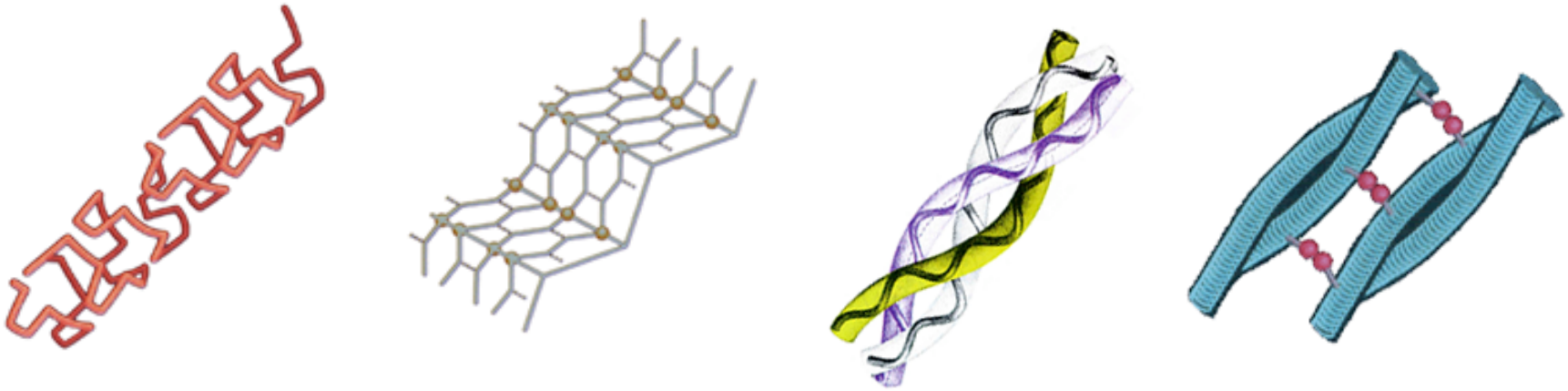


Figure 2. Natural evolutionary processes have afforded an array of structurally diverse protein-based materials that perform a variety of important biological functions.

synthesis.

Amphiphilic block copolymers

Block copolymers, which represent a unique category of synthetic polymers, consist of well-defined blocks where the individual blocks display different mechanical, chemical, or biological properties[13,14]. The synthetic copolymers, which undergo spontaneous self-assembly into ordered architectures, have been employed as polymer surfactants, pressure-sensitive adhesives, blend compatibilizers, thermoplastic elastomers, mineralization templates, and lithographic resists. Many conventional block copolymers of technological significance exhibit amphiphilicity. Amphiphilic block copolymers are composed of compositionally distinct blocks, which exhibit distinctly different interaction affinities for aqueous solutions[15,16]. These materials have attracted significant interest because of their complex phase behaviors in selective solvents, which mimic that of well-known amphiphiles like phospholipids. Amphiphilic diblock (**AB**) and triblock (**ABA** and **BAB**) copolymers are formed by two distinct polymer domains that undergo selective segregation of the hydrophobic domain in aqueous solvents to form well-defined mesoscopic aggregates. The aggregate architecture is characterized by a micellar structure in which the corona of the micelle is derived from the hydrophilic block (**A**) and the core of the micelle from the hydrophobic block (**B**). In the case of triblock copolymers, the resulting aggregate consists of a solvent-swollen network of spherical particles (micelles) derived from the insoluble endblock and linked together through the soluble midblock[17].

The phase behavior of amphiphilic block copolymers can be modified conveniently through manipulation of the macromolecular architecture, specifically, the length, composition, and sequence of the individual blocks[18]. Furthermore, the balance between the hydrophilic and hydrophobic character of the system can be modified through variation of the relative length of

the respective blocks. The amphiphilic block copolymers are attractive candidates for controlled delivery and release applications over conventional surfactant amphiphiles. In particular, amphiphilic block copolymers exhibit low critical micelle concentrations, high aggregate stabilities, slow unimer-micelle exchange rates, and a wide array of controlled aggregate sizes and morphologies. Because of the versatility of these materials, amphiphilic block copolymers are ideally suited for medical applications, such as emulsifiers, delivery agents, dispersants, gelation agents, compatibilizers, and foamants.

Many biomaterial scaffolds are derived from conventional organic monomers, including polymers of lactic and glycolic acids and their copolymers, which hydrolytically degrade to their biocompatible components. Polyethylene glycol-containing amphiphilic block copolymers are able to afford micelles as well as hydrogel materials. Hydrogels are water-swollen, cross-linked polymeric networks produced by the simple reaction of branched monomers or by association bonds such as hydrogen bonds and strong van der Waals interactions between chains[19]. Although amphiphilic block polymers derived from conventional organic monomers have technologically useful chemical and physical properties, they are limited in the number of block sequences that can be created. Polypeptides that comprise hydrophilic/lipophilic block structures may display the same technologically important physical characteristics, including the ability to reversibly self-assemble from aqueous solution to form an array of supramolecular aggregates with controlled microarchitectures. The assemblies derived from these amphiphilic block copolymers may be employed as biomaterials for *in vivo* applications due to their biocompatibility. The greater control of polypeptide microstructure, through design and selection of primary structure, may afford advantages over similar synthetic polymer systems currently studied. For these

studies, the elastin consensus repeat sequence (Val-Pro-Xaa-Yaa-Gly) has been employed as a substrate for the construction of synthetic polymers and block copolymers[2,17,20].

Elastin

Native elastin, a protein-based material, is a major component of the extracellular matrix and distributed widely throughout the body in tissues including the heart, blood vessels, and skin. Elastin confers rubber-like elasticity to the connective tissues, undergoing high deformation without rupture, storing energy involved in deformation, and then recovering to its original state once stress is removed[21]. The ability of elastin, along with other elastomeric proteins like abductin[22], tropoelastin[23-25], bysuss[26], and silk[27,28], to exhibit such resiliency relates to its primary and secondary structure, protein self-assembly, and other intermolecular forces that determine the formation of true and virtual cross-links. Specifically, elastin is characterized by two significant structural features inherent of all elastomeric materials. First, elastin is composed of monomers consisting of repetitive glycine-rich peptide motifs that are flexible and conformationally unconstrained. Secondly, elastin is cross-linked to form a network. The size and properties of the mobile monomer domain and the degree of intermolecular cross-linking can influence the mechanical behavior of the resulting elastin network.

Elastin is formed from a natural, soluble precursor tropoelastin. At the primary level, tropoelastin consists of alternating hydrophobic domains composed of mostly nonpolar amino acids and hydrophilic domains rich in lysine residues interspersed with alanine residues (Table 1)[29]. Within the hydrophobic domains are observed three to six peptide repeats with sequences such as GGVP, GVGVP, and GVGVAP. The monomers of tropoelastin are covalently cross-linked through the lysine residues that are dispersed throughout the protein to form hydrogel-like matrices. *In vivo* the elastin network is composed of two morphologically different components,

Table 1. 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 GGVP, GVGVP, 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.

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

an amorphous elastin constituent distributed within an organized microfibrillar, cross-linked scaffold primarily composed of fibrillin[29]. The hydrophobic domain is proposed to play a role in fiber formation through a coacervation mechanism, described for elastin by Urry and coworkers[30]. The second law of thermodynamics dictates that the entropy of a total system of constant composition must increase with increased temperature. In a multicomponent system, a component is said to have undergone an inverse temperature transition, T_t, when it experiences a decrease in entropy with an increase in temperature. Increasing the temperature of soluble elastin above the T_t results in a turbid solution due to the self-assembly and aggregation of the hydrophobic chains of elastin molecules. The resulting formation of elastin molecules that are held together by hydrophobic forces is known as a coacervate. The association of elastin molecules is thermodynamically controlled and thus, the coacervation phase is readily reversed upon cooling of the solution[30]. Elastin is an attractive target for the design of tissue-engineered analogues with potential application as medical biomaterials because of its elastomeric mechanical response and stability as well as the ability to form a hydrogel and undergo coacervation. The well-defined correlation between elastin sequence and macromolecular properties enables the construction of a wide array of biosynthetic analogues with tunable biophysical properties[31,32].

Elastin-mimetic polymers and block copolymers

The local secondary structure, macromolecular thermodynamics, and viscoelastic properties of the hydrophobic, elastomeric domains of native elastin can be emulated by elastin-mimetic polypeptides generated via concatenation of the consensus oligopeptide motifs. The most common of the sequence motifs is the pentapeptide (Val-Pro-Gly-Val-Gly)[20]. Polypeptides based on the pentameric repeat sequence undergo a reversible, temperature-dependent

hydrophobic assembly from aqueous solution analogous to the formation of a coacervate by native tropoelastin. The assembly of the elastin-mimetic polymer above the inverse temperature transition (T_t) is the result of spontaneous phase separation of the polypeptide, which coincides with a conformational rearrangement of the local secondary structure within the pentapeptide motifs. It has been shown that the material properties of the elastin-mimetic polymer are modulated through substitution of the amino acid in the third (Xaa) and fourth (Yaa) residues of the pentapeptide repeat sequence[20]. The Yaa position of the pentapeptide tolerates a wide variety of non-conservative amino acid substitutions. The T_t of the elastin-mimetic polypeptides can be altered through substitution at this position in a manner commensurate with the effect of the polarity of the amino acid side chain on the polymer-solvent interaction[31,32]. Addition of more hydrophobic residues lowers the value of T_t, while the addition of more polar residues increases the value of T_t (Figure 3)[33]. Also, replacement with Ala for the consensus Gly residue in third (Xaa) position of the pentapeptide repeat induces a change in the mechanical response of the material from elastic to plastic. This effect has been attributed to a change from a type II to a type I β turn arrangement at Pro-Gly.

Thus, elastin-mimetic blocks can be varied from elastomeric to plastic sequences to provide versatility in the mechanical properties of the synthetic construct. In addition, the transition temperature of the individual blocks can be modulated via substitution of amino acid residues in the fourth position of the pentapeptide with the desired polarity profile. The phase separation of hydrophobic domains within elastin-mimetic block copolymers can be exploited as a general mechanism for the reversible self-assembly of protein-based materials[2,17]. The combination of biocompatibility[34], physiological stability[20,35], high recombinant protein

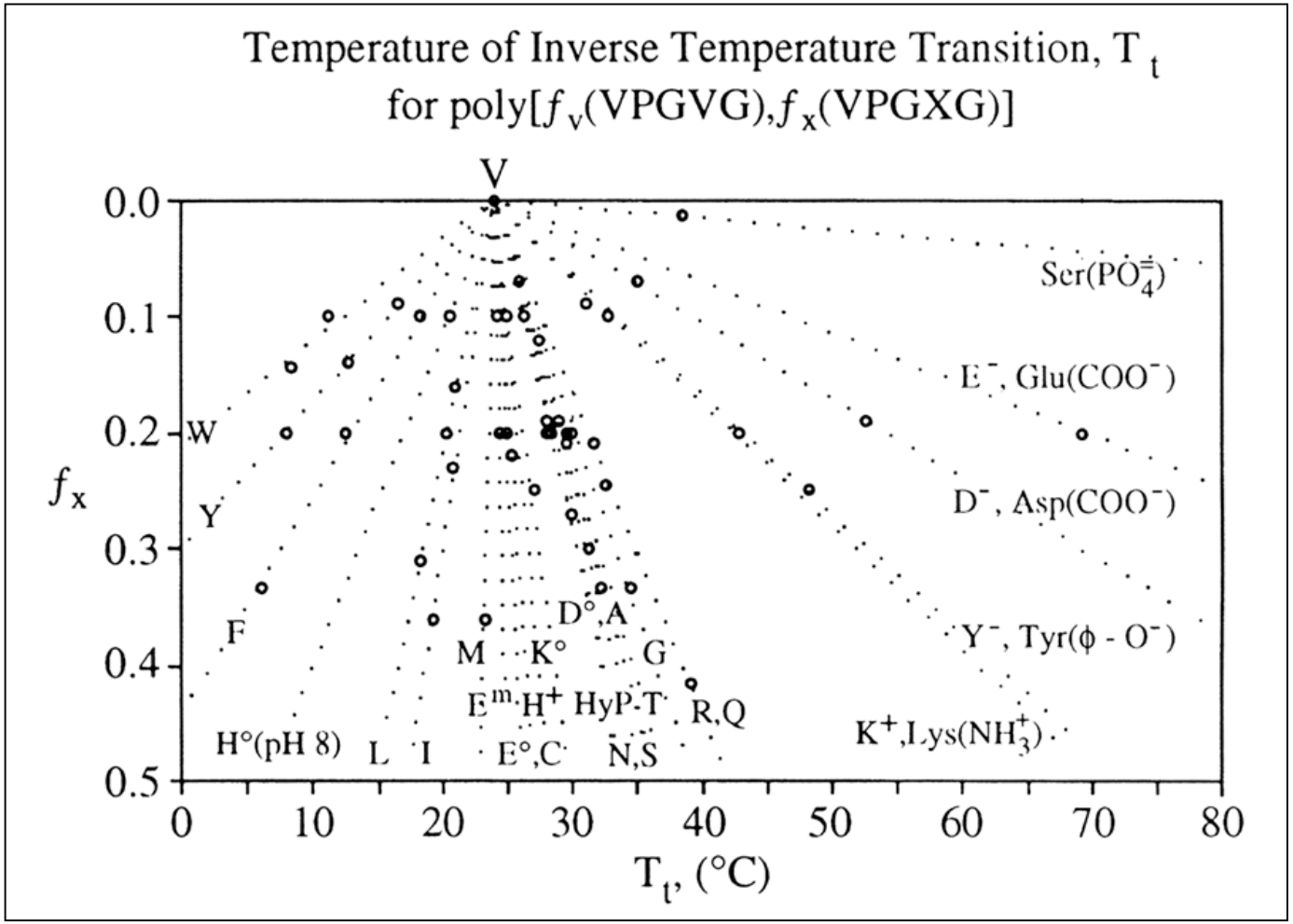


Figure 3. Characterization of inverse temperature transition, T_t, for elastin-mimetic polypeptides is shown graphically. Substitution of valine at position X with hydrophilic or hydrophobic amino acid residues causes a corresponding increase or a decrease in the transition temperature (T_t). In the hydrophobicity scale, f_v and f_x are the mole fractions of each peptide, where $f_v + f_x = 1$.

yields[36,37], and tunable responsive properties of elastin-mimetic polypeptides[31,32] makes protein-based materials derived these polymers and their block copolymers attractive candidates for the design and synthesis of novel biomaterials.

Previously, our lab has reported the synthesis of amphiphilic triblock (**BAB**-type) copolymers derived from elastin-mimetic peptides sequences[17]. The protein-based materials undergo reversible, temperature-dependent phase separation of the hydrophobic block in aqueous solution to form potentially biocompatible hydrogels under physiologically relevant conditions. The hydrophobic block (**B**) is derived from the plastic pentapeptide sequence [Val/Ile-Pro-Ala-Val-Gly] and has a predicted LCST below 37 °C. The hydrophilic block (**A**), which displays high values of T_t, is based on the elastomeric repeat sequence [(Val-Pro-Gly-Yaa-Gly)] where polar glutamic acid residues are introduced at the Yaa position periodically throughout the repetitive sequence. The change in polarity across the sequence results in a triblock (**BAB**-type) polypeptide that behaves as an amphiphilic block copolymer. Furthermore, the block sequences were chosen to have distinct mechanical properties such that the polypeptide's transition from aqueous solution to hydrogels above the T_t., emulates the behavior of conventional polymeric thermoplastic elastomers[3]. The triblock protein copolymer has the potential to act as a two-phase network where phase separation of the hydrophobic, plastic endblocks (**B**) occurs above their T_t, while the hydrophilic, elastic midblocks (**A**) remain hydrated under these conditions and act as virtual cross-links between the phase-separated, plastic domains. The reversible self-assembly of the polymer through segregation of the protein blocks into compositionally, structurally, and spatially distinct domains affords a hydrogel consisting of ordered structures on the mesoscopic scale. These properties of amphiphilic triblock (**BAB**-type) copolymers based on elastin-mimetic sequences have been verified through a combination of calorimetric methods, spectroscopic

analyses, and mechanical testing below and above the critical solution temperature[17,38,39]. Synthetic sequence-repetitive polypeptides, derived from elastin-mimetic sequences, have been prepared, which have the potential to emulate the biophysical properties of the native material. Through selection of amino acid sequence, the physical properties of these materials can be tailored to allow processing of the materials using technologies similar to those employed for conventional organic block copolymers.

The physically cross-linked protein-based materials possess a number of advantages over their chemically cross-linked counterparts, including ease of processing and the ability to avoid the addition or removal of chemical reagents or unreacted intermediates. In addition, the physical cross-linking strategy makes possible the incorporation of biologically or chemically active agents or cells that might otherwise be sensitive to covalent cross-linking schemes. Moreover, if the blocks are of sufficient size and chemical diversity, it is possible to access a diverse array of polymer morphologies. As a result, it is possible to tune a wide range of functional responses, such as mechanical behavior, permeability of drug elution characteristics, as well as the potential to design templated materials[14,40,41]. Prior studies of triblock (**BAB**-type) copolymers suggest that the density and strength of the physical cross-links are important determinants of both mechanical response and long-term material stability of two-phase protein networks[39]. In physically cross-linked hydrogels derived from elastin-mimetic polypeptides, the hydrophobic protein chains will entangle within the solvent swollen domain. In this respect, the extent of entanglements and their capacity to break and reform dynamically in response to stress will impact the mechanical properties. To investigate the effect of hydrophobic block content on the mechanical behavior of protein-based materials derived from elastin-mimetic triblock copolymers, a novel cloning strategy was developed that permitted construction of multi-block

copolymers with flexibility in selection of block architecture (e.g. molecular weight, sequence, and order) and seamless cloning between blocks. New polymers **B10** and **B11**, which contain hydrophobic endblocks nearly twice as large as previous variants of the triblock protein polymer, were synthesized in a manner that is both modular and convergent. This study describes the application of the new cloning strategy for biosynthesis of amphiphilic block copolymers derived from elastin-mimetic sequences. The amphiphilic triblock copolymers have the potential to form physically cross-linked polymer networks. Notwithstanding the desirable features of the physically cross-linked materials, these so-called *virtual* cross-links and the related assemblies of hydrophobic domains may be deformed or damaged at applied stresses lower than those required to disrupt covalent cross-links.

Cross-linkable elastin-mimetic polymers

As previously mentioned, elastin in its native state is formed through the enzymatic cross-linking of tropoelastin into an insoluble matrix upon deposition. After secretion, inter- and intra-molecular cross-linking is initiated by selective oxidative deamination of the lysine residues within the hydrophilic cross-linking domain by the copper-dependent enzyme lysyl oxidase[42,43]. The corresponding allysine undergoes further condensation with proximal allysine or lysine residues forming stable covalent cross-links including allysine aldol, lysinonorleucine, merodesmosine, and unique tetrafunctional cross-links such as desmosine and isodesmonsine[29,44]. In this way, the sequence of the polypeptide determines the positions and density of the cross-links, which occur at well-defined intervals along the polypeptide chain. The structural uniformity of these natural elastic materials is presumably responsible for their unique materials properties, i.e., high resilience and nearly ideal elastic behavior. The biosynthesis of native elastin networks serve as a paradigm for the design of elastin-mimetic materials that retain

the sequence definition and structural uniformity of their natural counterparts, while permitting the introduction of functional groups that facilitate formation of extended networks *under controllable conditions*.

Generally, cross-linking of elastin-mimetic protein polymers has been investigated using enzymatic, photoinitiated[45], or chemical approaches. For example, studies have investigated the application of transglutaminase or lysyl oxidase for enzymatic-based cross-linking[46]. Our lab has also explored solid-state cross-linking of recombinant elastin-mimetic proteins using both UV and visible light activated photoinitiators[21]. In most cases, however, recombinant elastin analogues have been prepared via cross-linking through available amino groups with chemical cross-linkers, including isocyanates, NHS-esters, phosphines, aldehydes, or genipin[40,47-56]. Similarly, our lab has described the design of an elastin-mimetic polypeptide $[(\text{Val-Pro-Gly-Val-Gly})_4(\text{Val-Pro-Gly-Lys-Gly})]_n$, in which lysine residues were chemically cross-linked using bis(sulfosuccinimidyl) suberate and disuccinimidyl suberate[57]. In addition, a glutaraldehyde cross-linking method has been applied to an amphiphilic (**BAB**) block copolymer interspersed with lysine-rich domains permitting both *physical* and *covalent* cross-linking[58]. While these approaches do provide a measure of control over the degree of cross-linking, in many cases, the chemical nature as well as the location of the cross-link is often ill defined. Limitations in chemical cross-linking via lysine residues include the cytotoxicity of the chemical agent, unintended side reactions, and residual amino groups, which may alter the responsive properties of the material and create non-cross-linked defect sites within the polymer network. In the studies described here, a new class of elastin-mimetic polypeptides were designed and synthesized, in which the fourth position of the consensus pentapeptide repeat unit has been substituted with a diverse set of unnatural amino acid analogues. The non-canonical amino acids

were selected for their unique and novel side chain functionalities that possess the ability to form covalent cross-links under *physiologically relevant conditions*.

Introduction of non-canonical amino acids

The engineering of proteins that contain non-canonical amino acids is an important development for investigating protein function, developing robust proteins, and incorporating building blocks with expanded technological utility for protein-based materials. While site-directed mutagenesis allows one to modify protein composition and function, the relatively limited scope of chemical functionality among the 20 natural amino acids provides significant opportunities for the chemist and chemical biologist. Recently, methods for incorporating non-canonical amino acid residues have advanced substantially and now constitute powerful and versatile tools for the redesign of natural proteins and *de novo* synthesis of protein analogues with materials applications[59]. In the 1950s, Cohen and coworkers laid the foundation for work on unnatural amino acids by demonstrating the substitution of methionine by selenomethionine in bacterial proteins[60]; a substitution, which many years later, revolutionized protein crystallography[61]. Protein chemists and engineers have now developed both chemisynthetic[62] and biosynthetic[63-65] approaches for incorporating hundreds of non-canonical amino acids into proteins. As a result, unnatural amino acids with an array of structural and electronic properties have extended our ability to manipulate the physiochemical and biological properties of engineered proteins.

For example, chemisynthetic strategies, including solid-phase synthesis[66,67] and expressed protein ligation[62], have been utilized to modify protein backbones[68], introduce fluorescent probes into peptides and proteins, and create modified signaling proteins, ion channels, and histones[69-71]. The use of these strategies, however, is characterized by limitations in selection of protein

sequence and length due to the need for protecting group chemistry and the restrictions on sites of chemical ligation. Biosynthetic approaches have also been developed for the *in vitro* and *in vivo* synthesis of proteins containing non-canonical amino acids. Generally, *in vitro* biosynthetic routes involve tRNAs, which recognize nonsense, frameshift, or coding codons, that are aminoacylated with the unnatural amino acid target and utilized in cell-tree translation systems. This method has been used for incorporation of amino acids with modified backbones[72,73], fluorophores[74-76], functional groups related to post-translational modifications[77], and photo- and chemically reactive side chains[78]. By microinjection or transfection of the aminoacylated tRNAs into living cells[79,80], unnatural amino acids spanning a range of electronic, structural, and conformational properties have been incorporated into proteins *in vivo*[81-83]. These biosynthetic strategies are limited by the synthesis and stability of the aminoacyl-tRNA adducts, the stoichiometry of aminoacylated tRNAs, which cannot be continuously generated, and the stress placed on cells by the microinjection and transfection techniques.

Another biosynthetic strategy, appropriate for living cells, takes advantage of the promiscuity of wild-type aminoacyl-tRNA synthetases (aaRS) to incorporate non-canonical analogues that are close structural analogues of the canonical amino acids[64,84]. In this approach, expression is carried out in a strain auxotrophic for one of the 20 canonical amino acids under selective pressure. Specifically, the cells are starved of the natural amino acid but supplemented with the analogue. The diversity of unnatural amino acids that can be incorporated via this strategy has been extended by overexpression of the wild-type synthetase[85] or a mutant synthetase generated through active-site[86,87] or editing domain[88] mutagenesis. Since the biosynthetic strategy results in global incorporation of amino acid analogues in response to the codon for the natural amino acid, it is commonly referred to as residue-specific incorporation.

Residue-specific incorporation allows for substantial alterations in the physical properties of proteins and enables chemical modification at multiple sites. Our lab has successfully used a residue-specific strategy for incorporation of a diverse set of proline analogues[89]. We have observed high degrees of biosynthetic substitution via overexpression of wild-type or mutant prolyl-tRNA synthetases and used analogue incorporation to investigate the role of proline in the macromolecular structure and function of biosynthetic polypeptides[90].

Lastly, the *in vivo* biosynthetic incorporation of unnatural amino acids can also be achieved site-specifically using an orthogonal tRNA/synthetase pair in response to "blank" codons (i.e. stop or frameshift codons) (Figure 4[91]). To ensure the success of this strategy, the components of the expression system must satisfy a number of criteria. First, the unnatural amino acid must have good cellular bioavailability and be metabolically stable. In addition the analogue and corresponding tRNA should be compatible with the elongation factor and ribosome, but should not be substrates for the host strain synthetases. Second, the blank codon must be recognized by the orthogonal tRNA, but not any endogenous tRNAs. Third, the orthogonal tRNA/synthetase pair must be specific for the unnatural amino acid, function in the host strain, and exhibit no cross-reactivity with any endogenous tRNA/synthetase pairs in the host organism (i.e. function orthogonally)[92]. In 1998, Furter first demonstrated a successful *in vivo* strategy for the site-specific incorporation of non-canonical amino acids through the use of an heterologous tRNA/synthetase pair from yeast in an *E. coli* expression host[93]. Schultz and coworkers have made important contributions to the development of this strategy. Specifically, high-throughput selection schemes[94] enable the evolution of aminoacyl-tRNA synthetases that show high specificity toward structurally dissimilar unnatural amino acid analogues with technologically useful functional groups, including acetyl[95] and benzophenone moieties[96].

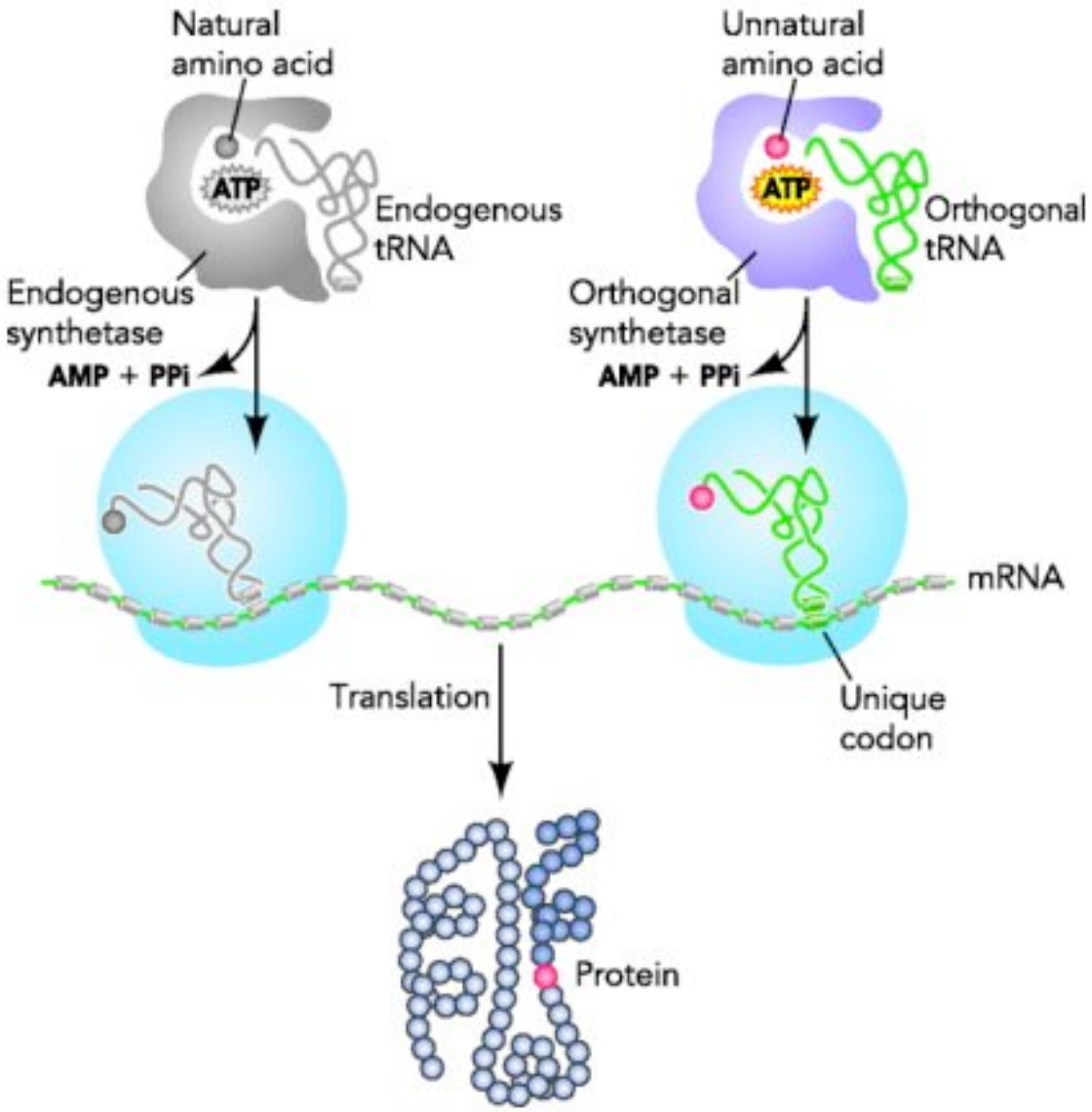


Figure 4. The genetic encoding of unnatural amino acids (UAA) during *in vivo* protein synthesis utilizes a new set of components consisting of an orthogonal tRNA, an orthogonal aminoacyl-tRNA synthetase, and a unique codon. The orthogonal tRNA is acylated with the UAA specifically by the orthogonal synthetase. During protein translation, the acylated tRNA delivers the UAA in response to a "blank" codon that does not encode for any of the 20 canonical amino acids

In this study, the residue-specific incorporation strategy is used for global incorporation of non-canonical amino acids, which display close structural similarity to the canonical amino acid and useful side chain functionality. As a result, we have generated elastin-mimetic polypeptides with the potential to be cross-linked or to form covalent conjugates to other molecules under physiologically relevant conditions. The residue-specific strategy is well-suited for the generation of cross-linked elastin analogues since the strategy can be highly efficient and used to prepare proteins in relatively high yields, as required for application in materials science[59]. In addition, we report the incorporation of a structurally dissimilar, photo-cross-linkable analogue of tyrosine via an amber suppression strategy using an orthogonal tRNA/synthetase pair. By means of a modified *E. coli* host strain, the site-specific strategy is extended to incorporate the unnatural amino acid at multiple sites periodically within an elastin-mimetic polypeptide with high specificity.

References

[1] Van Hest, J. C. M.; Tirrell, D. A. *Chemical Communications* **2001**, *4*, 1897-1904.

[2] Lee, T. A. T.; Cooper, A.; Apkarian, R. P.; Conticello, V. P. *Advanced Materials* **2000**, *12*, 1105-1110.

[3] Noshay, A.; Mcgrath, J. E. *Block copolymers: overview and critical survey*; Academic: New York, 1977.

[4] Banta, S.; Wheeldon, I. R.; Blenner, M. *Annual Review of Biomedical Engineering* **2010**, *12*, 167-186.

[5] Payne, S. C.; Patterson, M.; Conticello, V. P. In *Protein Engineering Handbook*; Lutz, S., Bornscheuer, U. T., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2009; Vol. 1, p 915-936.

[6] Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Biotechnology Progress* **1990**, *6*, 198-202.

[7] Goldberg, I.; Salerno, A. J.; Patterson, T.; Williams, J. I. *Gene* **1989**, *80*, 305-314.

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

[9] Chilkoti, A.; Christensen, T.; MacKay, J. A. *Current Opinion in Chemical Biology* **2006**, *10*, 652-657.

[10] Herrero-Vanrell, R.; Rincón, A. C.; Alonso, M.; Reboto, V.; Molina-Martinez, I. T.; Rodríguez-Cabello, J. C. *Journal of Controlled Release* **2005**, *102*, 113-122.

[11] Hoffman, S.; Knecht, S.; Langer, R.; Kaplan, D. L.; Vunjak-Novakovic, G.; Merkle, H. P.; Meinel, L. *Tissue Engineering* **2006**, *12*, 2729-2738.

[12] Langer, R.; Tirrell, D. A. *Nature* **2004**, *428*, 487-492.

[13] Ferrari, F. A.; Cappello, J. In *Protein-based materials*; McGrath, K. P., Kaplan, D. L., Eds.; Birkhäuser: Boston, 1997, p 37-60.

[14] Petka, W. A.; Hardin, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A. *Science* **1998**, *281*, 389-392.

[15] Alexandridis, P. *Current Opinion in Colloid and Interface Science* **1996**, *1*, 490-501.

[16] Mortenson, K. *Current Opinion in Colloid and Interface Science* **1998**, *3*, 12-19.

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

[18] Hajduk, D. A.; Kossuth, M. B.; Hillmyer, M. A.; Bates, F. S. *Journal of Physical Chemistry B* **1998**, *102*, 4269-4276.

[19] Peppas, N. A. In *Biomaterials science: an introduction to materials in medicine*; Ratner, B. D., Hoffman, A. S., Schoen, F. J., Lemons, J. E., Eds.; Academic Press: New York, N.Y., 1996, p 100-106.

[20] Wright, E. R.; Conticello, V. P. *Advanced Drug Delivery Reviews* **2002**, *54*, 1057-1073.

[21] Nagapudi, K.; Brinkman, W. T.; Leisen, J. E.; Huang, L.; McMillan, R. A.; Apkarian, R. P.; Conticello, V. P.; Chaikof, E. L. *Macromolecules* **2002**, *35*, 1730-1737.

[22] Cao, Q.; Wang, Y.; Bayley, H. *Current Biology* **1997**, *7*, R677-R678.

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

[24] Sandberg, L. B.; Soskel, N. T.; Leslie, J. G. *New England Journal of Medicine* **1981**, *304*, 566-579.

[25] Urry, D. W.; Luan, C.-H.; Harris, C. M.; Parker, T. M. In *Protein-based materials*; McGrath, K. P., Kaplan, D. L., Eds.; Birkhäuser: Boston, 1997, p 133-178.

[26] Deming, T. J. *Current Opinion in Chemical Biology* **1999**, *3*, 100-105.

[27] Hayashi, C. Y.; Lewis, R. V. *Journal of Molecular Biology* **1998**, *275*, 773-784.

[28] Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V. *International Journal of Biological Macromolecules* **1999**, *24*, 271-275.

[29] Vrhovski, B.; Weiss, A. S. *European Journal of Biochemistry* **1998**, *258*, 1-18.

[30] Urry, D. W. *Journal of Protein Chemistry* **1988**, *7*, 1-34.

[31] Urry, D. W.; Gowda, D. C.; Parker, T. M.; Luan, C.-H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D. *Biopolymers* **1992**, *32*, 1243-1250.

[32] Urry, D. W.; Luan, C.-H.; Parker, T. M.; Gowda, D. C.; Prasad, K. U.; Reid, M. C.; Safavy, A. *Journal of the American Chemical Society* **1991**, *113*, 4346-4347.

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

[34] Urry, D. W.; Parker, T. M.; Reid, M. C.; Gowda, D. C. *Journal of Bioactive and Compatible Polymers* **1991**, *6*, 263-282.

[35] Mecham, R. P.; Broekelman, T. J.; Filiszar, C. J.; Shapiro, S. D.; Welgus, H. G.; Senior, R. M. *Journal of Biological Chemistry* **1997**, *272*, 18071-18076.

[36] Chow, D. C.; Dreher, M. R.; Trabbic-Carlson, K.; Chilkoti, A. *Biotechnology Progress* **2006**, *22*, 638-646.

[37] Daniell, H.; Guda, C.; McPherson, D. T.; Zhang, X.; Xu, J.; Urry, D. W. *Methods in Molecular Biology* **1997**, *63*, 359-371.

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

[39] Wu, X. Y.; Sallach, R. E.; Haller, C. A.; Caves, J. A.; Nagapudi, K.; Conticello, V. P.; Levenston, M. E.; Chaikof, E. L. *Biomacromolecules* **2005**, *6*, 3037-3044.

[40] Bellingham, C. M.; Lillie, M. A.; Gosline, J. M.; Wright, G. M.; Starcher, B. C.; Bailey, A. J.; Woodhouse, K. A.; Keeley, F. W. *Biopolymers* **2003**, *70*, 445-455.

[41] Welsh, E. R.; Tirrell, D. A. *Biomacromolecules* **2000**, *1*, 23-30.

[42] Bressan, G. M.; Prockop, D. J. *Biochemistry* **1977**, *16*, 1406-1412.

[43] Kagan, H. M.; Sullivan, K. A. *Methods in Enzymology* **1982**, *82*, 637-650.

[44] Reiser, K.; McCormick, R. J.; Rucker, R. B. *Faseb Journal* **1992**, *6*, 2439-2449.

[45] Fancy, D. A.; Kodadeck, T. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 6020-6024.

[46] McHale, M. K.; Setton, L. A.; Chilkoti, A. *Tissue Engineering* **2005**, *11*, 1768-1779.

[47] Vieth, S.; Bellingham, C. M.; Keeley, F. W.; Hodge, S. M.; Rousseau, D. *Biopolymers* **2007**, *85*, 199-206.

[48] Girotti, A.; Reguera, J.; Rodríguez-Cabello, J. C.; Arias, F. J.; Alonso, M.; Ma Testera, A. *Journal of Materials Science: Materials in Medicine* **2004**, *15*, 479-484.

[49] Martino, M.; Tamburro, A. M. *Biopolymers* **2001**, *59*, 29-37.

[50] Lee, J.; Macoscko, C. W.; Urry, D. W. *Biomacromolecules* **2001**, *2*, 170-179.

[51] Nowatzki, P. J.; Tirrell, D. A. *Biomaterials* **2004**, *25*, 1261-1267.

[52] Trabbic-Carlson, K.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2003**, *4*, 572-580.

[53] McMillan, R. A.; Lee, T. A. T.; Conticello, V. P. *Macromolecules* **1999**, *32*, 3643-3648.

[54] Zio, K. D.; Tirrell, D. A. *Macromolecules* **2003**, *36*, 1553-1558.

[55] Lim, D. W.; Nettles, D. L.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2003**, *9*, 222-230.

[56] Lim, D. W.; Nettles, D. L.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2007**, *8*, 1463-1470.

[57] McMillan, R. A.; Conticello, V. P. *Macromolecules* **2000**, *33*, 4809-4821.

[58] Sallach, R. E.; Cui, W.; Wen, J.; Martinez, A.; Conticello, V. P.; Chaikof, E. L. *Biomaterials* **2009**, *30*, 409-422.

[59] Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. *Current Opinion in Chemical Biology* **2010**, *14*, 774-780.

[60] Murnier, R.; Cohen, G. N. *Biochimica et Biophysica Acta* **1956**, *21*, 592-593.

[61] Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. *EMBO Journal* **1990**, *9*, 1655-1672.

[62] Muir, T. W. *Annual Review of Biochemistry* **2003**, *72*, 249-289.

[63] Budisa, N.; Minks, C.; Alefelder, S.; Wenger, W.; Dong, F.; Moroder, L.; Huber, R. *Faseb Journal* **1999**, *13*, 41-51.

[64] Link, A. J.; Mock, M. L.; Tirrell, D. A. *Current Opinion in Biotechnology* **2003**, *14*, 603-609.

[65] Wang, L.; Schultz, P. G. *Chemical Communications* **2002**, 1-11.

[66] Kimmerlin, T.; Seebach, D. *Journal of Peptide Research* **2005**, *65*, 229-260.

[67] Merrifield, R. B. *Advances in Enzymology and Related Areas of Molecular Biology* **1969**, *32*, 221-296.

[68] Baca, M.; Kent, S. B. H. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 11638-11642.

[69] Flavell, R. R.; Muir, T. W. *Accounts of Chemical Research* **2009**, *42*, 107-116.

[70] Valiyaveetil, F. I.; Leonetti, M.; Muir, T. W.; MacKinnon, R. *Science* **2006**, *314*, 1004-1007.

[71] Vazquez, M. E.; Nitz, M.; Stehn, J.; Yaffe, M. B.; Imperiali, B. *Journal of the American Chemical Society* **2003**, *125*, 10150-10151.

[72] Eisenhauer, B. M.; Hecht, S. M. *Biochemistry* **2002**, *41*, 11472-11478.

[73] Koh, J. T.; Cornish, V. W.; Shultz, P. G. *Biochemistry* **1997**, *36*, 11314-11322.

[74] Murakami, H.; Hohsaka, T.; Ashizuka, Y.; Hashimoto, K.; Sisido, M. *Biomacromolecules* **2000**, *1*, 118-125.

[75] Taki, M.; Hohsaka, T.; Murakami, H.; Taira, K.; Sisido, M. *FEBS Letters* **2001**, *507*, 35-38.

[76] Taki, M.; Hohsaka, T.; Murakami, H.; Taira, K.; Sisido, M. *Journal of the American Chemical Society* **2002**, *124*, 14586-14590.

[77] Fahmi, N. E.; Dedkova, L.; Wang, B.; Golovine, S.; Hecht, S. M. *Journal of the American Chemical Society* **2007**, *129*, 3586-3597.

[78] Kanamori, T.; Nishikawa, S.-I.; Shin, I.; Schultz, P. G.; Endo, T. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 485-490.

[79] Köhrer, C.; Yoo, J.-H.; Bennett, M.; Schaack, J.; RajBhandary, U. L. *Chemistry and Biology* **2003**, *10*, 1095-1102.

[80] Rodriguez, E. A.; Lester, H. A.; Dougherty, D. A. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 8650-8655.

[81] Beene, D. L.; Dougherty, D. A.; Lester, H. A. *Current Opinion in Neurobiology* **2003**, *13*, 264-270.

[82] Dougherty, D. A. *Current Opinion in Chemical Biology* **2000**, *4*, 645-652.

[83] England, P. M.; Zhang, Y.; Dougherty, D. A.; Lester, H. A. *Cell* **1999**, *96*, 89-98.

[84] Hendrickson, T. L.; de Crécy-Lagard, V.; Schimmel, P. *Annual Review of Biochemistry* **2004**, *73*, 147-176.

[85] Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. *Angewandte Chemie International Edition* **2000**, *39*, 2148-2152.

[86] Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *Journal of the American Chemical Society* **2002**, *124*, 5652-5653.

[87] Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS Letters* **2000**, *467*, 37-40.

[88] Döring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crécy-Lagard, V.; Schimmel, P.; Marlière, P. *Science* **2001**, *292*, 510-504.

[89] Kim, W.; George, A.; Evans, M.; Conticello, V. P. *ChemBioChem* **2004**, *5*, 928-936.

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

[91] Lin, M. Z.; Wang, L. *Physiology* **2008**, *23*, 131-141.

[92] Liu, D. R.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 4780-4785.

[93] Furter, R. *Protein Science* **1998**, *7*, 419-426.

[94] Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

[95] Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 56-61.

[96] Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 11020-11024.

Chapter 2

Generation of amphiphilic block copolymers derived from elastin-mimetic sequences

Adapted in part from Payne, S. C.; Patterson, M. A.; Conticello, V. P. In *Protein Engineering Handbook*, Lutz, S., Bornscheuer, U. T., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2009; Vol. 1, p 915-936.

Introduction

The production of protein-based materials derived from polypeptides of repeating sequence is best accomplished using the techniques of recombinant DNA (rDNA) technology and protein expression. The synthesis of the protein is dictated at the DNA level allowing for high-level protein expression in a bacterial host[1,2]. Target proteins are generated using an *in vivo* biosynthesis strategy that allows for production of polypeptides with predetermined DNA and subsequent amino acid sequences. This approach is advantageous over a conventional chemical synthetic strategy where oligopeptides are limited to low degrees of polymerization (< 60 residues) due to decreasing yield with each successive coupling step[3]. Genetic engineering methods are able to directly produce, with high fidelity, synthetic polypeptides of exact amino acid sequence and high molecular weight. A general scheme for the biological synthesis of repetitive polypeptides involves chemical synthesis of the corresponding DNA coding sequence, ligation of the DNA cassette into a plasmid vector, propagation in a bacterial host, and finally, inducible expression of the polypeptide (Figure 1b[4]).

Within the context of these studies, a 'protein-based material' refers to a sequence-repetitive polypeptide or a multi-domain polymer, where one or more individual domains consists of a sequence-repetitive polypeptide that is encoded within a synthetic DNA expression cassette. The biosynthesis of these high molecular weight polymers usually requires the construction of large, synthetic genes composed of tandem repeats of 'monomeric' DNA. Automated DNA synthesis technology is currently limited to the production of oligodeoxynucleotides of about a hundred base pairs in length. Therefore, direct synthesis of the entire gene cannot be employed to generate medium to high molecular weight polypeptides In addition, such highly repetitive DNA sequences may be susceptible to homologous

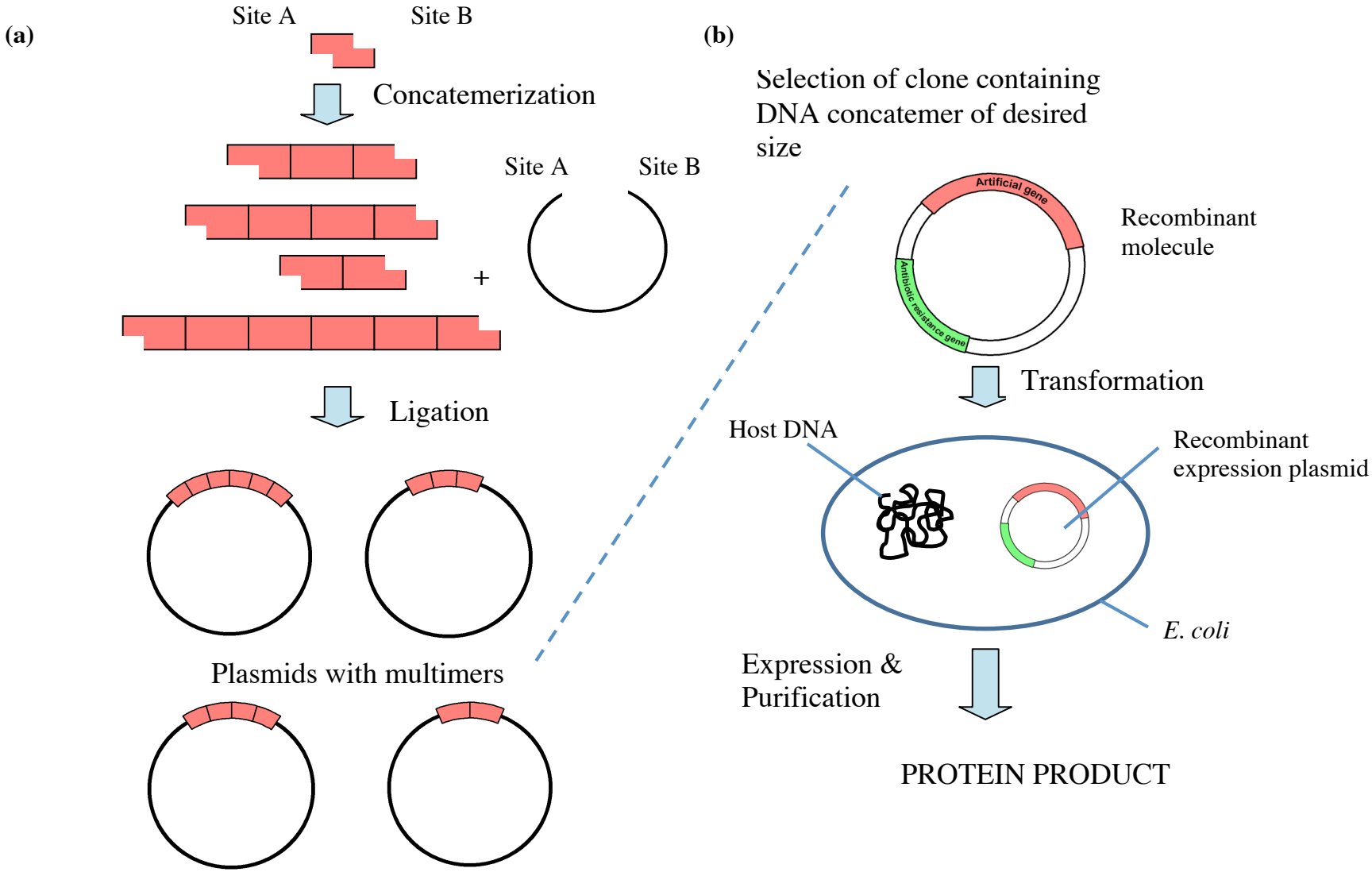


Figure 1. (a) Representation of DNA cassette concatemerization, a commonly applied method for the synthesis of concatemeric genes encoding sequence repetitive polypeptides. (b) General strategy for biological synthesis of polymers for protein-based materials.

recombination and the plasmid clones may be unstable *in vivo*. Thus, the cloning and expression of repetitive genes requires unique considerations beyond conventional DNA synthesis routes[5-7]. A recent review has comprehensively described the range of experimental methods that have been developed for the synthesis of repetitive genes encoding sequence-repetitive polypeptides[8]. Our lab has successfully utilized DNA cassette concatemerization (Figure 1a[9]) to generate repetitive oligonucleotide sequences encoding polymers and block copolymers consisting of sequence-repetitive polypeptides[10-16].

DNA cassette concatemerization begins with the synthesis of double-stranded, oligonucleotide fragments known as DNA 'monomers', which contain non-palindromic cohesive ends[17-20]. Generation of the 'sticky'-ended DNA monomers is typically accomplished through the use of restriction endonucleases that recognize and cleave non-palindromic sequences. The length of one oligopeptide repeat is selected such that it can be encoded by an oligonucleotide of approximately 50 to 150 base pairs, which is readily generated by automated DNA synthesis. Self-ligation of the DNA monomers proceeds in a head-to-tail fashion to generate a library of concatemers that differ in length by one DNA monomer unit. The DNA concatemers are separated by agarose gel electrophoresis according to the degree of concatemerization. Concatemers of desired size range are then extracted from the gel and enzymatically inserted into the cloning plasmid. The size of individual concatemers are identified by screening the population of clones either using colony screening polymerase chain reaction (PCR) or restriction digestion of purified recombinant plasmids.

Construction of protein-based, amphiphilic block copolymers has employed DNA cassette concatemerization to generate synthetic genes encoding sequence-repetitive polypeptides derived from elastin-mimetic sequences[14,16]. The hydrophilic and hydrophobic

blocks are encoded within separate DNA cassettes, concatemerized independently, and joined together via sequential ligation into a multifunctional polylinker to create genetic fusions of elastin-mimetic polymers. A cloning strategy depicting preparation of the synthetic DNA cassette encoding a previously studied elastin-mimetic triblock copolymer **B9**[16] is shown in Figure 2[9]. Briefly, the double-stranded DNA monomers **E** and **P**, which encode the elastic-like and plastic-like repeat sequences, respectively, are synthesized. Cleavage of **E** and **P** by *Sex*A I and *Bsp*M I correspondingly generates non-palindromic 5'-cohesive ends with single-stranded extensions, which ensure exclusive head-to-tail concatemerization of the DNA monomers. As a type IIs restriction endonuclease, *Bsp*M I cleaves the DNA strand downstream of the recognition sequence. Therefore, the sequence of the resulting single-stranded extensions depends solely on the position of the cleavage site and can be made non-palindromic through the choice of the downstream DNA sequence. Two related expression plasmids were constructed with polylinkers that allowed for insertion of the plastic concatemers, which were subsequently joined together to form an acceptor plasmid for the elastic concatemer library. The elastic concatemers were then inserted into the compatible *Sex*A I site to afford a series of expression constructs that encoded triblock copolymers containing central blocks of varying length. The original approach described has several disadvantages. First, a time-consuming series of cloning steps was required to prepare the expression construct encoding the triblock polypeptide. Second, while the size of the elastic midblock could be easily varied using this strategy, the plastic endblock domains could not be modified unless the entire expression cassette was reconstructed. In addition, the peptide sequences at the block termini or junctions could not be altered using this approach.

In order to facilitate the construction of amphiphilic triblock copolymers with plastic endblocks of longer chain length, a more versatile design scheme was devised that took

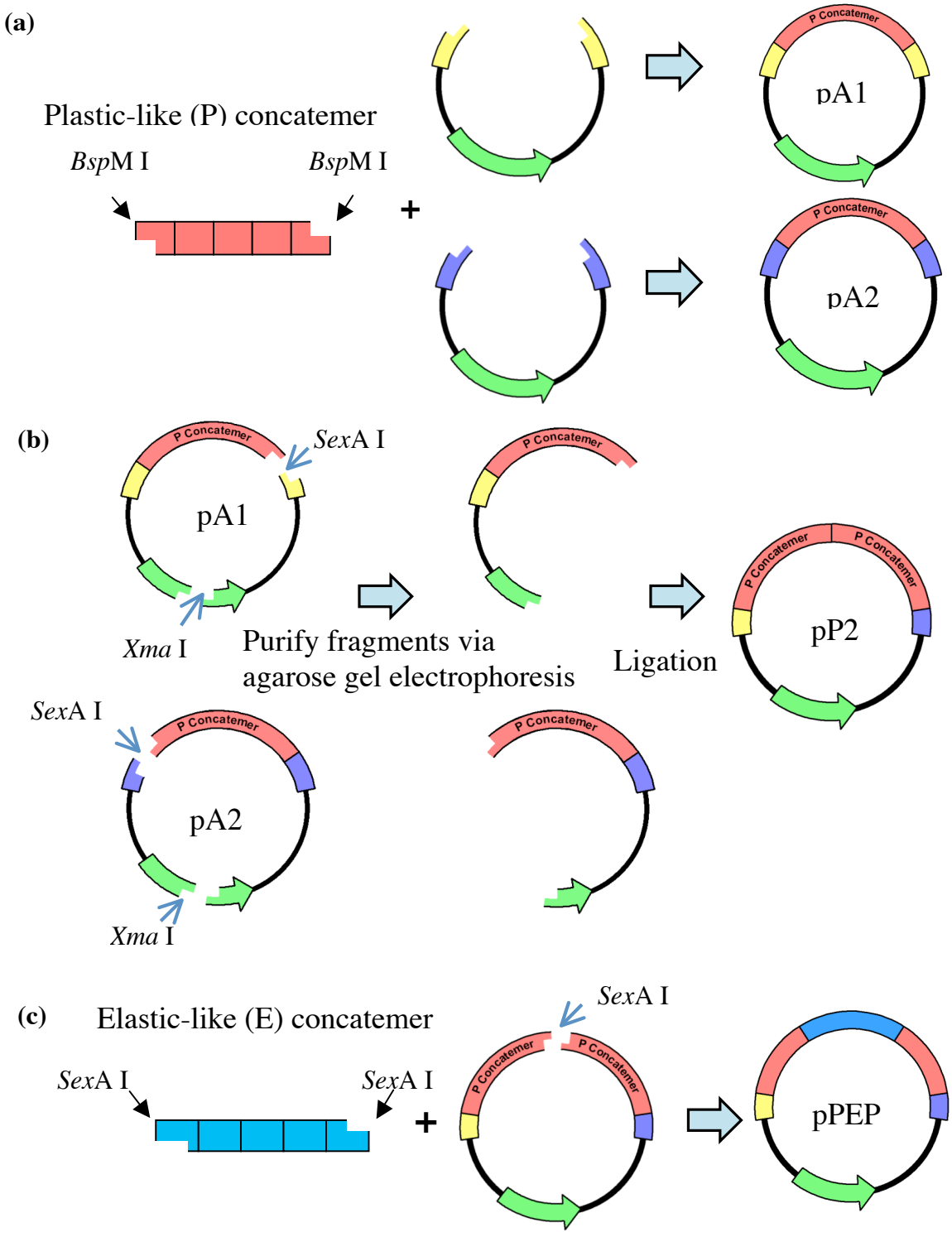


Figure 2. Previously reported strategy for the assembly of a triblock polypeptide. (a) DNA concatemers encoding plastic cassettes were inserted into two modified expression vectors, which afforded two plasmids **pA1** and **pA2**. (b)These plasmids were subsequently joined together to form an acceptor plasmid, **pP2**, for insertion of the elastic concatemer library. (c) Strategy for assembly of the triblock polypeptide through insertion of the elastic concatemer library into the compatible *Sex*A I site of the acceptor plasmid.

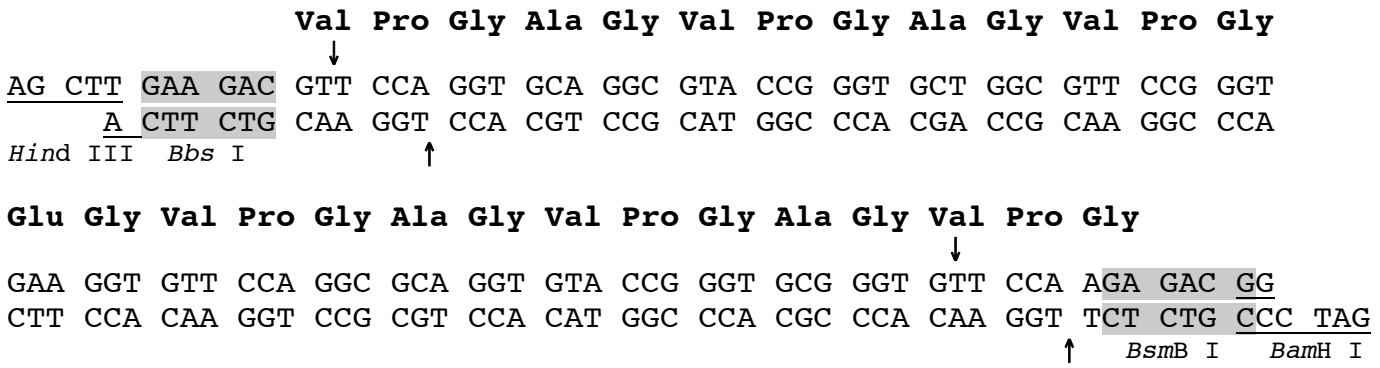
advantage of a seamless cloning technique[21]. The seamless cloning strategy utilizes the recognition and cleavage characteristics of the type IIs restriction endonucleases, in which cleavage of a DNA duplex occurs at a specific site downstream of its recognition site. Cleavage of the DNA 'monomer' usually generates synthetic duplexes with 5'-cohesive ends in which the sequence of the bases within the overhangs is independent of the recognition site. The position of the sequence can be defined to facilitate directional cloning without the need for endonucleases with unique recognition and cleavage patterns that match the internal sequence of the template DNA duplex. Furthermore, the restriction sites for type IIs endonucleases are cleaved from the DNA cassette during the cloning procedure and thus, eliminated from the coding sequence of the DNA monomer.

In the seamless cloning strategy, type IIs restriction endonucleases can be employed to generate DNA monomers with non-palindromic, complementary cohesive ends that are joined together in an exclusively head-to-tail manner to construct a library of concatemers. The technique can also be utilized for the directional assembly of concatemers into blocks of defined size and sequence in a way that is both modular and convergent. Since the multiple cloning sites of common commercially available expression vectors do not contain compatible restriction endonuclease cleavage sites, a modified expression plasmid is constructed via insertion a polylinker (Scheme 2) containing type IIs restriction sites. In the expression of the amphiphilic block copolymers derived from elastin-mimetic sequences, our lab has employed the pET series of plasmids. Following cloning of the triblock DNA concatemer gene fusion into the modified expression plasmid, transcription of the amphiphilic copolymer gene originates downstream of a T7 lac promoter in response to induction of a coliphage T7 RNA polymerase gene. In this case, the polymerase is provided by the host strain BL21-Gold (DE3), a lambda lysogen carrying the

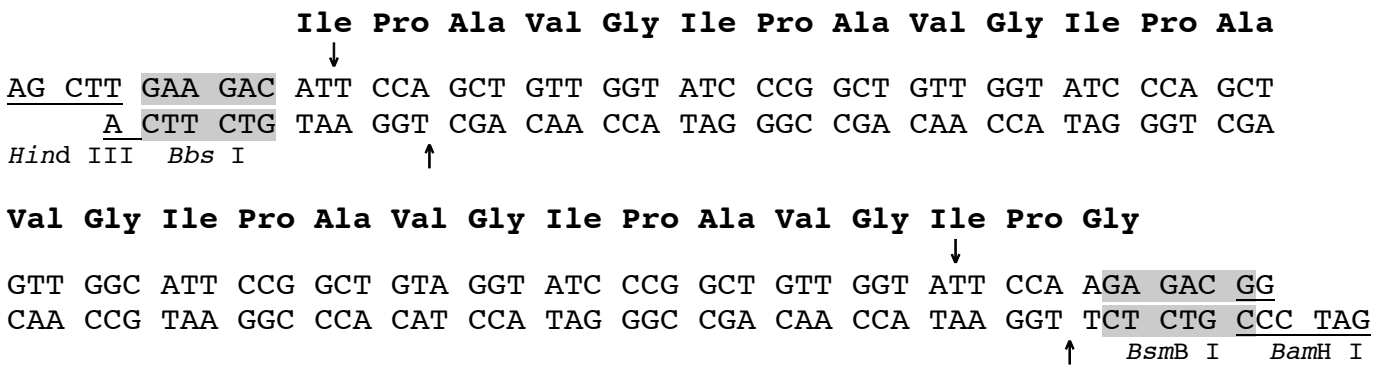
T7 RNA polymerase gene under control of the *lac*UV5 promoter. The target amphiphilic block copolymers are generated via expression in Terrific Broth (TB) media according to a hyperexpression protocol, where dilution of the repressor protein in rapidly growing cells permits binding of the RNA polymerase and initiation of transcription without addition of the inducer IPTG[22]. The elastin-mimetic polymers are then purified from the cell lysate via repetitive cycling through the inverse temperature of transition, T_t, where precipitation of the target protein occurs above the T_t in the presence of a high concentration of salt[23]. The biosynthesis and purification methods allow for the production of up to gram quantities of the desired polypeptide[12,13,16,24,25].

The seamless cloning strategy has been has been employed to construct both elastic (**S1**) and plastic (**S3**) concatemers (Scheme 1), which are derived from the elastin-mimetic pentapeptide repeat sequence (Val-Pro-Xaa-Yaa-Gly). The newly reported synthesis technique should facilitate cloning of multi-domain polymers with improved flexibility in selection of domain architecture (e.g. size, sequence, order) and result in seamless cloning between blocks. The **BAB**-type triblock copolymers should exhibit the technologically useful properties of elastin-mimetic polypeptides, including a T_t and undergo coacervation. Furthermore, the amphiphilic polymers should undergo reversible microscopic phase separation from aqueous solution to potentially form thermoplastic elastomer hydrogels above the T_t. The physically cross-linked networks derived from the new elastin-mimetic triblock copolymers should exhibit improved mechanical properties due to the increase in hydrophobic block content.

(a)



(b)



Scheme 1. (a,b) Oligonucleotide cassettes encoding the elastic- **S1** (a) and plastic-like **S3** (b) repeat sequences. The recognition sites for the relevant restriction endonucleases, which are employed for generation of the DNA monomer, are highlighted. Arrows indicate the cleavage positions on the sense and anti-sense strand for the respective type IIs restriction endonucleases.

(a)

Sequence of primers used for construction of expression plasmid adaptor

Elastin Adaptor-F

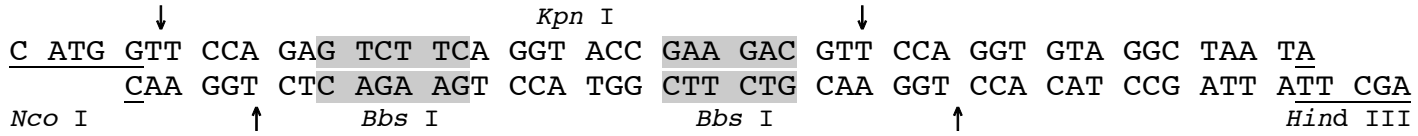
5′-CATGGTTCCAGAGTCTTCAGGTACCGAAGACGTTCCAGGTGTAGGCTAATA-3′

Elastin Adaptor-R

5′-AGCTTATTAGCCTACACCTGGAACGTCTTCGGTACCTGAAGACTCTGGAAC-3′

(b)

Adaptor sequence after annealing



Scheme 2. (a) Sequence of **Elastin Adaptor** forward and reverse primers before and after annealing is shown. (b) Oligonucleotide cassette encoding the **Elastin Adaptor** is depicted following annealing. The recognition sites for the relevant restriction endonucleases, which are employed for cloning of the triblock copolymer expression plasmid, are highlighted. Arrows indicate the cleavage positions on the sense and anti-sense strand for the respective endonucleases.

Experimental Methods

Materials

All chemical reagents were purchased from Fisher Scientific, Inc. (Pittsburgh, PA), Sigma-Aldrich Corp. (St. Louis, MO) or VWR International, LLC (Radnor, PA) unless otherwise noted. Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases were purchased from both New England Biolabs, Inc. (NEB, Beverly, MA) and Promega Corp. (Fitchburg, WI), while antartic phosphatase, T4 DNA ligase, and T4 polynucleotide kinase were purchased only from NEB. Shrimp alkaline phosphatase was obtained from Roche Applied Science (Indianapolis, IN). Plasmid pZErO®-2.1, pBAD/His A and DNA DipStick™ were obtained from Invitrogen Corp. (Carlsbad, CA). Pellet Paint® co-precipitant and plasmid pET-24 d (+) were purchased from Novagen (a brand of EMD Chemicals Inc., Gibbstown, NJ). Plasmid DNA preparation and purification were performed using the Mini-Prep 24™ (MacConnell Research Corporation, San Diego, CA), the QIAfilter Plasmid Maxi Kit, QIAprep Spin Miniprep Kit, QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA), and QIAwell 8 Ultra Plasmid Kit, and the DNA Clean & Concentrator™ and Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA). Synthetic oligonucleotides were purchased from either Sigma-Genosys (a division of Sigma-Aldrich Corporation, St. Louis, MO) or Integrated DNA Technologies (Coralville, IA) and were used as received. The *E. coli* strain, TOP10F´ was obtained from Invitrogen Corp. (Carlsbad, CA) while XL10-Gold ultracompetent and BL21-Gold (DE3) cells were purchased from Stratagene (La Jolla, CA). Plasmids designated pSCP were previously synthesized by Dr. Sonha C. Payne, a former member of the Conticello lab at Emory University.

General Methods

The molecular biology techniques utilized here, including expression cloning, polymerase chain reaction, propagation of bacteria, and gel electrophoresis were adapted from a standard molecular cloning manual[26] or the manufacturer protocol, unless otherwise described in detail. All reagents for the manipulation of DNA, bacteria, and proteins were sterilized by either autoclave or passage through a syringe filter (0.2 μm cellulose membrane) or vacuum filter unit (standard polyethersulfone (PES) membrane) available from VWR International, LLC (Radnor, PA). Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. The concentration of DNA solutions was measured by an Ultrospec 3000 UV/Visible Spectrophotometer (Pharmacia Biotech, Cambridge, UK). Polymerase chain reaction (PCR) was carried out using a GeneAmp 2400 Thermal Cycler (PerkinElmer Inc., Waltham, MA) or MJ Mini™ Gradient Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules, CA). Agarose gel electrophoresis images were captured using a Kodak DC-120 digital camera or Kodak Gel Logic 112 imaging system from Carestream Health, Inc. (Rochester, NY). Over the course of these studies, automated DNA sequencing was performed at each of the following facilities: Emory University Core DNA Sequencing Facility on a Perkin-Elmer ABI Prism 377 DNA sequencer and Beckman Coulter Genomics (Danvers, MA) and GENEWIZ, Inc. (South Plainfield, NJ) on an Applied Biosystems 3730xl DNA analyzer. Protein electrophoresis was performed using the Mini-PROTEAN® 3 Cell electrophoresis system from Bio-Rad Laboratories, Inc. (Hercules, CA) and the Prestained Protein Marker, Broad Range (10-175 kDa) from NEB (Beverly, MA). Proteins were visualized using a conventional negative staining protocol with zinc sulfate[27].

DNA cassette concatemerization

The synthesis of the elastic- and plastic-like DNA concatemers has been previously reported[9] and the procedure is described below. The single-stranded oligonucleotides corresponding to the sense and anti-sense strand of the DNA monomers **S1** and **S3** (Scheme 1) were chemically synthesized and annealed upon receipt to produce duplex monomer DNA. Oligonucleotide primers were dissolved in buffer EB (10 mM Tris-HCl, pH 8.5; QIAGEN Inc., Valencia, CA) to a final concentration of 0.5 μg/μL. A 20 μL aliquot (10 μg) of each of the sense and anti-sense primers were mixed together with 4 μL 5 M NaCl, 4 μL 1 M $MgCl_2$, and 152 μL sterile ddH_2O. The DNA strands were annealed together by gradually decreasing the temperature of the reactions from 99 °C to 30 ºC, by 1 °C every 5 min, and then maintaining a temperature of 30 ºC for 15 min. Synthesis of the DNA duplex was visualized by agarose gel electrophoresis (4% NuSieve® GTG® agarose; 1X Tris-Borate-EDTA (TBE) buffer). Ethanol precipitation with Pellet Paint® co-precipitant was used to purify the double-stranded DNA fragments. Addition of 5'-phosphates to the double-stranded fragments was performed with T4 polynucleotide kinase (10 μL; 1000 units) by incubating solutions for 2 h at 37 °C (50 μL DNA, 75 μL sterile ddH_2O, 15 μL T4 DNA ligase buffer with ATP). The reactions were extracted with phenol/chloroform/isoamyl alcohol (25:24:1) to remove enzyme and then ethanol precipitation was used to isolate the phosphorylated DNA duplexes and reduce the salt concentration. Approximately 350-500 ng of each phosphorylated, DNA monomer was isolated and stored in 50 μL buffer EB at -20 °C.

Ligations between the synthetic DNA monomers (100 ng) and a *Bam*H I/*Hin*d III-digested pZErO®-2.1 plasmid were performed in a 3:1 molar ratio in the presence of 2 μL T4 DNA ligase enzyme (800 units) and 1X T4 DNA ligase buffer in a final reaction volume of 20

μL at 16 ºC for 12 h. Aliquots (2 μL) of these ligation mixtures were used to transform chemically competent cells of *E. coli* strain TOP10F´ (50 μL). The cells were then recovered in 800 μL pre-warmed SOC rich media at 37 °C for 1 h with shaking at 225 rpm. Aliquots (150 μL) of each transformation mixture were spread onto low salt Luria Broth (LB) agar media supplemented with 1 mM IPTG and kanamycin (50 μg/mL) for antibiotic selection. The plates were incubated at 37 °C for 12 to 14 h. Eight single colonies were picked from the plate and used to inoculate separate culture tubes of low salt LB media (5 mL) supplemented with kanamycin, which were grown overnight at 37 ºC in a rotating drum. The plasmid DNA was isolated using the QIAwell 8 Ultra Plasmid Kit for each of the samples. Recombinant plasmids were screened via PCR amplification of the monomer insert by primers designed to bind the M13 reverse and M13 (-20) forward priming sites located upstream and downstream of the target sequence, respectively (Appendix 1). PCR products were analyzed by agarose gel electrophoresis (4% NuSieve® GTG agarose®, 1X TBE buffer) for the presence of the DNA monomer insert. Automated DNA sequencing analysis with the M13 primers confirmed the sequences of S1 and S3 monomer genes. Recombinant plasmids containing the correct insert sequences of DNA monomers **S1** and **S3** flanked by *Bbs* I and *Bsm*B I restriction endonuclease recognition sites were designated **pSCP1** and **pSCP2**, respectively, and stored at –20 ºC when not in use.

Because DNA cassette concatemerization required a large amount of DNA monomer, plasmid DNA was prepared from four-2.8 L flasks containing 500 mL of low salt LB media supplemented with kanamycin (50 μg/mL). Approximately 0.5 to 1.5 mg of **pSCP1** and **pSCP2** plasmid DNA was isolated using the QIAfilter Plasmid Maxi Kit. The plasmid DNA (~500 μg of each) was digested with *Bbs* I (1 unit/μg) in 1X NEBuffer 2 for 16 h at 37 ºC. The *Bbs* I-digested DNA fragments were purified by extraction with phenol/chloroform/isoamyl alcohol (25:24:1)

and isolated via ethanol precipitation. The linearized plasmid DNA was then digested with *Bsm*B I (1.5 units/μg) in 1X NEBuffer 3 at 55 °C for 16 h. The *Bbs* I/*Bsm*B I-cut **S1** and **S3** DNA monomers were purified independently by extraction with phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol precipitation. Each of the digested DNA monomers was loaded onto a preparative 2% NuSieve® GTG® agarose gel (1X TBE buffer) in order to separate the pZErO®-2.1 vector fragment from the monomer insert. Bands corresponding to 75 bp (monomers **S1** and **S3**) were 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 yielding approximately 5-6 μg of monomer.

A ligation between monomers (multimerization) was prepared by incubating separate solutions of each isolated DNA monomer in the presence of 5 μL 1X T4 DNA ligase buffer and 1 μL T4 DNA ligase (400 units) at 16 °C for 16 h. The multimerization products from each reaction were separated by agarose gel electrophoresis (1% agarose gel, 0.5X TBE buffer), which revealed a ladder of DNA multimers (Figure 4) differing in size by a single monomer unit (75 bp). Concatemers between 1000 and 3000 bp were excised from the gel, purified using the Zymoclean™ Gel DNA Recovery Kit and quantified using DNA DipStick™ (approximately 20 ng DNA in 21 μL buffer EB). Acceptor plasmids were prepared by digesting the **pSCP1** and **pSCP2** monomer plasmids (20 μg) with only *Bbs* I (2 units/μg) at 37 ° for 4 h. The digested plasmids were extracted with phenol/chloroform/isoamyl alcohol (25:24:1) and purified by ethanol precipitation. The *Bbs* I-digested monomer acceptor plasmids were dephosphorylated at the 5'-ends with shrimp alkaline phosphatase (1 unit/pmol) in 1X shrimp alkaline phosphatase (SAP) buffer at 37 °C for 30 min. The dephosphorylated acceptor plasmids were isolated via 1% agarose gel electrophoresis (0.5X TBE buffer) and purified using the Zymoclean™ Gel DNA

Recovery Kit. The purified **S1** and **S3** DNA concatemers of size 1000 to 3000 bp were then ligated back into the corresponding *Bbs* I-digested **pSCP1** or **pSCP2** monomer acceptor plasmids. Fifty nanograms of each insert and vector were ligated together in a 1:1 ratio (w/w) with 1 μL T4 DNA ligase enzyme and 1X T4 DNA ligase buffer in a 10 μL reaction volume at 16 °C for 16 h. Chemically competent cells of TOP10F´ (50 μL) were transformed with the ligation mixtures (5 μL) and spread onto low salt LB agar media supplemented with 1 mM IPTG and kanamycin (50 μg/mL). The plates were incubated at 37 °C for 12 to 14 h. Twenty-four colonies were picked from each ligation plate and used to inoculate separate culture tubes (2 mL each) of Magnificent Broth (MacConnell Research Corporation, San Diego, CA) supplemented with kanamycin. The cell cultures were incubated at 37 ºC with shaking at 230 rpm for 20 h. The plasmid DNA was purified using the Mini-Prep 24^{TM}. Recombinant plasmids were screened by double digestion with *Bam*H I/*Hin*d III. Digestion products were analyzed by agarose gel electrophoresis (1% agarose; 0.5X TBE buffer) for the presence of the multimer inserts.

From the pool of transformants, four colonies of interest (two from **S1** and two from **S3** DNA concatemer ligations) were isolated containing a multimer in the size range of 1500 to 2500 bp. The recombinant plasmids were used to transform chemically competent TOP10F' cells and purified from low salt LB media (500 mL) supplemented with kanamycin using the QIAfilter Plasmid Maxi Kit. Next, plasmids encoding amphiphilic block copolymers were prepared using the **S1** and **S3** DNA concatemer plasmids, **pSCP3.1** and **pSCP3.11** as well as **pSCP4.7** and **pSCP4.11**, respectively, as starting material. It is worth noting that the recombinant plasmid variants contain concatemer genes of different length. Specifically, the **pSCP3.1** and **pSCP3.11** contain multimers of approximately 2000 and 1500 bp, respectively, while **pSCP4.7** and **pSCP4.11** contain multimers corresponding to approximately 2500 and

2200 bp in length (Figure 6).

Assembly of DNA concatemer cassettes into triblock gene fusions

Assembly of the triblock gene fusions first required construction of two diblock assemblies (one of each of **AB**- and **BA**-type). The protocol utilized in our lab by Payne and coworkers[9] for assembly of the diblock plasmids is described below. To generate a plasmid encoding a diblock copolymer with the hydrophilic, elastic block (**S1**) at the N-terminus and the hydrophobic, plastic block (**S3**) at the C-terminus (Figure 7), the plasmids **pSCP3.1** and **pSCP4.7** (10 μg) were digested first with *Nco* I (5 units/μg) in 1X NEBuffer 4 at 37 °C for 4 h. In addition, a plasmid encoding a second diblock copolymer with the hydrophobic plastic block (**S3**) at the N-terminus and the hydrophilic elastic block (**S1**) at the C-terminus was constructed by digesting **pSCP4.11** and **pSCP3.11** (10 μg) with *Nco* I. All *Nco* I-digested mixtures were purified with the QIAquick PCR Purification Kit using a microcentrifuge. The *Nco* I-cut **pSCP3.1** and **pSCP4.11** plasmids were then digested with *Bsm*B I (5 units/μg) in 1X NEBuffer 3 at 55 °C for 4h, while the *Nco* I-cut **pSCP4.7** and **pSCP3.11** plasmids were digested with *Bbs* I (5 units/μg) in 1X NEBuffer 2 at 37 °C for 4h. The digestion mixtures were separated via agarose gel electrophoresis (0.75% SeaPlaque®GTG® agarose, 0.5X TBE buffer). Each of the **S1** and **S3** DNA concatemer-containing plasmid fragments were excised from the gel and purified via the Zymoclean™ Gel DNA Recovery Kit with a yield of approximately 1.8 to 4.9 μg DNA.

Two ligation reactions were performed by incubating the **pSCP3.1** and **pSCP4.7** as well as the **pSCP4.11** and **pSCP3.11** (100 ng each) pairs of DNA fragments with 1 μL T4 DNA ligase (400 units) and 1X T4 DNA ligase buffer in a reaction volume of 20 μL at 16 °C for 16 h. An aliquot (2 μL) of each ligation mixture was used to transform electrocompetent cells of *E. coli* strain TOP10F′ (40 μL). The cells were recovered in 1 mL SOC rich media for 1 h at 37 °C

with shaking at 225 rpm. Aliquots (150 μL) of each transformation mixture were spread onto low salt LB agar media supplemented with 1 mM IPTG and kanamycin (50 μg/mL). The plates were incubated at 37 °C for 12-14 h. Twenty-four colonies were picked from each plate and used to inoculate separate culture tubes of low salt LB media (5 mL) supplemented with kanamycin, which were incubated overnight at 37 ºC on a rotating drum. The plasmid DNA was isolated using the QIAwell 8 Ultra Plasmid Kit for each of the samples. Recombinant plasmids were screened by double digestion with *Bam*H I/*Hin*d III. Digestion products were analyzed by agarose gel electrophoresis (0.75% SeaPlaque®GTG® agarose, 0.5X TBE buffer) for the presence of the diblock inserts.

From the pool of transformants, five plasmids containing an insert size concomitant with a gene fusion of the **S1** and **S3** DNA concatemers and two plasmids containing an **S3** and **S1** diblock insert were used to transform electrocompetent TOP10F' cells and purified from low salt LB media (500 mL) supplemented with kanamycin using the QIAfilter Plasmid Maxi Kit (approximately 44 μg plasmid DNA isolated). The sequence of the purified plasmid DNA was examined by automated DNA sequencing analysis with M13 primers for the presence of each diblock flanked by *Bbs* I and *Bsm*B I restriction endonuclease recognition sites at the 5'- and 3'-end of the coding strand, respectively. Recombinant plasmids containing the correct sequence of the **S1** to **S3** DNA concatemer gene fusion were designated **pSCP5.15** and **pSCP5.24**, while the correctly sequenced clones of the **S3** to **S1** DNA concatemer gene fusion were designated **pSCP6**. The recombinant diblock clones were stored at –20 ºC when not in use. It is important to note that the diblock gene fusions within the **pSCP5** variants are composed of **S1** and **S3** DNA concatemers of unique length.

Plasmids encoding triblock copolymers containing hydrophilic elastin midblocks (**S1**)

flanked by hydrophobic plastin endblocks (**S3**) were then prepared using the seamless cloning technique (Figure 8). The plasmid **pSCP4.11** (10 μg) and the diblock plasmids **pSCP5.15**, **pSCP5.24** and **pSCP6** were digested first with *Nco* I (5 units/μg) and purified with the QIAquick PCR Purification Kit using a microcentrifuge. The *Nco* I-cut **pSCP4.11** and **pSCP6** plasmids were then digested with *Bsm*B I, while the *Nco* I-cut **pSCP5.15** and **pSCP5.24** plasmids were digested with *Bbs* I. The digestion mixtures were separated via agarose gel electrophoresis (0.75% SeaPlaque®GTG® agarose, 0.5X TBE buffer). Approximately, 3.2 to 6.3 μg of each of the plasmid fragments were excised from the gel and purified via the Zymoclean™ Gel DNA Recovery Kit.

To generate two triblock gene fusions, the pairs of **pSCP4.11** and **pSCP5.15** DNA fragments and **pSCP6** and **pSCP5.24** DNA fragments (100 ng each) were incubated separately with 1 μL T4 DNA ligase (400 units) and 1X T4 DNA ligase buffer in a 20 μL reaction volume at 16 °C for 16 h. Electrocompetent cells of TOP10F´ (40 μL) were transformed with the ligation mixtures (2 μL) and recovered in 1 mL SOC at 37 °C with shaking at 225 rpm for 1 h. Aliquots (150 μL) of each transformation mixture were spread onto low salt LB agar media supplemented with 1 mM IPTG and kanamycin (50 μg/mL). The plates were incubated at 37 °C for 12-14 h. Eight colonies were picked from each plate and used to inoculate separate culture tubes of low salt LB media (5 mL) supplemented with kanamycin grown overnight at 37 ºC in a rotating drum. The plasmid DNA was isolated using the QIAwell 8 Ultra Plasmid Kit for each of the samples. Recombinant plasmids were screened by digestion with *Bam*H I and *Hin*d III. Digestion products were analyzed by agarose gel electrophoresis (0.75% SeaPlaque®GTG® agarose, 0.5X TBE buffer) for the presence of the triblock inserts.

From the pool of transformants, five plasmids containing an insert size concomitant with a triblock gene fusion containing a single elastic (**S1**) concatemer midblock and three plasmids containing an elastic midblock twice as large were designated **pMAP3** and **pMAP10** (Figure 10), respectively and stored at –20 ºC when not in use. The resulting **pMAP3** and **pMAP10** plasmids contain DNA concatemer gene fusions of the hydrophobic (**S1**) and hydrophilic (**S3**) concatemers, which encode the corresponding triblock (**BAB**-type) copolymers **B10** and **B11**. The sequence of the purified plasmid DNA was examined by automated DNA sequencing analysis with M13 primers for the presence of the triblock flanked by *Bbs* I and *Bsm*B I restriction endonuclease recognition sites at the 5'- and 3'-end of the coding strand, respectively. Plasmids **pMAP3** and **pMAP10** were used to transform electrocompetent TOP10F' (40 μL) cells and XL10-Gold Ultracompetent cells, according the manufacturer protocol, the latter of which are suitable for cloning of large plasmids. Aliquots (150 μL) of each TOP10F' and XL10-Gold Ultracompetent transformation mixture were spread onto low salt LB agar media supplemented with 1 mM IPTG and kanamycin (50 μg/mL) or kanamycin, chloramphenicol (35 μg/mL) and tetracycline (10 μg/mL), correspondingly. The plates were incubated at 37 °C for 12 to 14 h. Single colonies of the **pMAP3** and **pMAP10** transformed TOP10F' and XL10-Gold cells were used to inoculate low salt LB media (5 mL) supplemented with kanamycin alone or kanamycin and chloramphenicol. The cells were cultured at 37 °C in a rotating drum to an optical density (measured at a wavelength of 600 nm) of 0.8 to 1.0. Glycerol stocks of the cell strains were prepared from 800 μL of cell culture and 200 μL of sterile 80% glycerol and stored at –80 °C.

Construction of the expression vector

The triblock gene fusions were inserted into a commercially available pET vector using a

two-step cloning procedure. Initially, the triblock gene fusions were cloned into a modified pBAD/His A vector that contained an adaptor sequence within the multiple cloning site. The single-stranded oligonucleotides corresponding to the sense and anti-sense strands of the DNA adapter **Elastin Adaptor** (Scheme 2) were chemically synthesized and annealed upon receipt to produce duplex monomer DNA. The *Nco* I and *Hin*d III-digested and dephosphorylated plasmid pBAD/His A (100 ng) was incubated with the **Elastin Adaptor** cassette in a 1: 3 molar ratio with T4 DNA ligase (1 μL, 400 units) and 1X T4 DNA ligase buffer in a reaction of 20 μL at 16 °C for 14 h. An aliquot (5 μL) of the ligation was used to transform electrocompetent cells of *E. coli* strain TOP10F′ (40 μL), which were plated onto LB agar media supplemented with ampicillin (100 μg/mL). The plates were incubated at 37 °C for 12 to 14 h. The plasmid DNA was isolated from 24 positive transformants using the QIAwell 8 Ultra Plasmid Kit. Recombinant plasmids were screened by double digestion with *Bbs* I/*Hin*c II. Digestion products were analyzed by agarose gel electrophoresis (4% NuSieve® GTG® agarose, 1X TBE buffer) to identify recombinant plasmids with the adaptor inserted. Clones with correct insert sizes were analyzed by automated DNA sequencing using the pBAD forward and reverse sequencing primers (Appendix 1). Two recombinant clones, designated **pSCP7**, containing the correct adapter sequence were used to transform electrocompetent TOP10F' cells and purified from LB media (500 mL) supplemented with ampicillin (100 μg/mL) using the QIAfilter Plasmid Maxi Kit (approximately ~30 to 40 μg plasmid DNA).

To generate the complimentary sticky ends necessary for seamless cloning of the triblock DNA concatemer gene fusions into the adaptor sequence of the expression plasmid, **pSCP7** (10 μg) was digested with *Bbs* I (200 units) in 1X NEBuffer 2 at 37 °C for 4 h. The digestion mixture was purified with the QIAquick PCR Purification Kit using a microcentrifuge. The *Bbs* I-cut

DNA was also digested with *Kpn* I to reduce the incidence of false positives during subsequent ligations. The digestion fragments were isolated via agarose gel electrophoresis (1% agarose, 0.5X TBE). The plasmid was excised from the gel and purified using the Zymoclean™ Gel DNA Recovery Kit. The linearized plasmid DNA (~1 μg) was dephosphorylated with shrimp alkaline phosphatase (10 units) by incubation for 30 min at 37 °C with 1X SAP buffer (10 μL). The digestion fragments were isolated via agarose gel electrophoresis (1% agarose, 0.5X TBE). The plasmid was excised from the gel and purified using the Zymoclean™ Gel DNA Recovery Kit.

The triblock elastin gene fusions were isolated via sequential digestion of plasmids **pMAP3** and **pMA10** with *Bbs* I and *Bsm*B I. The purified triblock inserts (100 ng) were incubated with the *Bbs* I/*Kpn* I-digested expression plasmid **pSCP7** (50 ng plasmid) at 16 °C for 16 h in 1 μL T4 DNA ligase enzyme (400 units) and 1X T4 DNA ligase buffer for a final reaction volume of 20 μL. Aliquots (2 μL) of each of the ligation reactions were used to transform electrocompetent TOP10F' cells, which were spread onto LB plates containing ampicillin (100 μg/mL) and incubated at 37 °C for 12-14 h. Eight single colonies of each of the transformants were used to inoculate LB media (5 mL) supplemented with ampicillin and cultured overnight at 37 °C in a rotating drum. The plasmid DNA was isolated and purified using the QIAprep Spin Miniprep Kit. Recombinant clones were screened for the presence of the triblock inserts by double digestion with *Nco* I and *Hin*d III. The digestion fragments were separated via agarose gel electrophoresis (1% agarose, 0.5X TBE) to identify recombinant plasmids. From the pool of transformants, five recombinant plasmids, designated **pMAP5** and **pMAP11**, containing an insert of size corresponding to the triblock gene fusions encoding **B10** and **B11**, respectively, were isolated. Glycerol stocks of the **pMAP5** and **pMAP11** TOP10F' transformants were prepared from 800 μL of overnight LB (5 mL) cell culture and 200 μL 80%

glycerol and stored at –80 °C.

The triblock gene fusions were removed from the pBAD/His A plasmid as DNA cassettes flanked by the **Elastin Adaptor**. Plasmids **pMAP5** and **pMAP11** (10 μg of each) were digested with *Nco* I (90 units), *Hin*d III (60 units), and *Hin*c II (50 units) in 1XNEBuffer 2 and 1X BSA at 37 °C for 4 h. The triblock inserts were separated via agarose gel electrophoresis (0.75% SeaPlaque®GTG® agarose, 0.5X TBE) and purified using the Zymoclean™ Gel DNA Recovery Kit (0.5 μg yield of DNA triblock cassettes). The purified elastin inserts (~120 ng) were then ligated into *Nco* I/*Hin*d III-digested pET-24 d(+) expression plasmid (~60 ng) via incubation with 1 μL T4 DNA ligase enzyme (400 units) and 1 μL T4 DNA ligase buffer at 16 °C for 16 h. Electrocompetent cells of TOP10F' were transformed with the ligation reactions (2 μL), and the recovery mixture was spread onto LB plates containing kanamycin (50 μg/mL). The plates were incubated at 37 °C for 12 to 14 h. Eight single colonies of the TOP10F' transformants were used to inoculate LB media (5 mL) supplemented with kanamycin and cultured overnight at 37 °C in a rotating drum. The plasmid DNA was isolated and purified using the QIAprep Spin Miniprep Kit. Recombinant clones were screened for the presence of the triblock insert by double digestion with *Nco* I and *Hin*d III. The digestion fragments were separated via agarose gel electrophoresis (0.75% SeaPlaque®GTG® agarose, 0.5X TBE) to identify recombinant plasmids. From the pool of transformants, ten recombinant plasmids (five of each), designated **pMAP6** and **pMAP12** (Figure 12), containing an insert of size corresponding to the triblock genes encoding amphiphilic triblock copolymers **B10** and **B11**, respectively, were isolated. The sequences of the inserts were confirmed by automated DNA sequencing analysis using the T7 promoter, forward primer and T7 terminator, reverse primer (Appendix 1). Plasmids **pMAP6** and **pMAP12** (1 μL) were used to transform fresh TOP10F' electrocompetent cells and XL10-Gold Ultracompetent

cells, according the manufacturer protocol. Glycerol stocks of the recombinant clones were prepared from 800 μL of overnight LB (5 mL) culture and 200 μL of sterile 80% glycerol and stored at –80 °C.

Expression and purification of elastin-mimetic block copolymers

Electrocompetent cells of *E. coli* strain BL21-Gold (DE3) were transformed with the unique recombinant plasmids **pMAP6** and **pMAP12** (1 μL each) encoding amphiphilic triblock copolymers, **B10** and **B11**, respectively. Aliquots of the recovery mixture (100 μL) were spread onto LB plates containing kanamycin (50 μg/mL) and tetracycline (10 μg/mL) and incubated at 37 °C overnight. The cells were subcultured in LB media (5 mL) supplemented with antibiotics at 37 °C in a rotating drum to an OD_{600nm} of 0.8 to 1.0. Glycerol stocks of the expression strains were prepared from 800 μL of cell culture and 200 μL of sterile 80 % glycerol and stored at –80 °C.

In addition, single colonies of positive transformants were used to inoculate 20 mL of Terrific Broth (TB) medium (12 g tryptone, 24 g yeast extract, 4 mL glycerol, 2.31 g KH_2PO_4, 12.54 g K_2HPO_4, 1 L ddH_2O) supplemented with kanamycin and tetracycline. The cultures were incubated at 37 °C shaking at 225 rpm overnight. The cell cultures were centrifuged at 4 °C, 4000 x *g* for 20 min, the supernatant decanted and the pellet suspended in 20 mL fresh TB media lacking antibiotics. The centrifuge step was repeated and 20 mL (2% of final volume) of suspended culture was used to inoculate 1 L fresh TB media divided between 2-2.8L flasks (490 mL each) and supplemented with the antibiotic kanamycin (30 μg/mL) at a reduced concentration to maximized protein expression. The cultures were grown for 36 h at 37 °C with agitation at 175 rpm.

Aliquots (1 μL) from each culture were removed at 24 and 36 h for further analysis by

SDS PAGE. The aliquots were centrifuged at 4 °C, 6300 rpm for 5 min. The supernatant was discarded and the cell pellet was suspended in 30 μL sterile ddH_2O and stored at –20 °C. Samples for SDS PAGE were prepared by mixing 5 μL of the whole cell lysate with 12.5 μL 2X SDS-gel loading buffer (100 mM Tris-Cl, pH 6.8; 4% (w/v) SDS, electrophoresis grade; 0.2% (v/v) bromophenol blue; 20% glycerol), 2.5 μL 1 M DTT, and sterile ddH_2O for a final volume of 25 μL. Samples were boiled at 100 °C for 10 min then cooled on ice (0 °C) for 10 min prior to gel loading. Following 36 h incubation, the remaining 1 L of cells were isolated via centrifugation at 4 °C, 4000 x *g* for 20 min.

The cell pellets from each 500 mL expression culture were suspended in 25 mL lysis buffer (50 mM sodium phosphate, 300 mM NaCl, pH 7.0). The suspended cell pellets from each 500 mL **B10** and **B11** culture were combined for a total volume of 50 mL. The cells were lysed via three freeze (-80 °C) / thaw (25 °C) cycles. Following the third thaw cycle, lysozyme (10 mg/mL stock solution) was added to a final concentration of 1 mg/mL and 5 mL protease inhibitor cocktail from Sigma-Aldrich Corp. (St. Louis, MO), prepared as follows 215 mg in 4 mL sterile ddH_2O and 1mL DMSO, was added to the lysate. The mixtures were incubated at room temperature (25 °C) for 30 min with agitation at 225 rpm. To the cell lysate, 1 M $MgCl_2$ was added to a final concentration of 1 mM and Benzonase® nuclease (Sigma-Aldrich Corp.; St. Louis, MO) to a final concentration of 25 U/mL. The cell lysate was incubated at 4 °C with agitation at 225 rpm for 36 hr. The cell lysate was centrifuged at 20,000 x *g* for 40 min at 4 °C to pellet the cellular debris.

The elastin-mimetic polymers were then purified via inverse transition cycling (ITC) according to a protocol adapted from Meyer and Chilkoti[28]. Five molar NaCl (~25 mL) was added to the 50 mL of soluble extracts for a final molarity of 2 M. The solutions were incubated

at 37 °C for 60 min with gentle shaking until precipitation of the elastin mimetic-polymers was observed as evidenced by increased turbidity of the solution. The solutions were centrifuged at 37 °C, 10,000 x *g* for 15 min to pellet the precipitated proteins ('hot-spin'). The supernatant was discarded, and the protein pellet was placed on ice. Ice-cold (0 °C), sterile lysis buffer (~42 mL total) plus 1 mL protease inhibitor cocktail was used to suspend the protein pellets. The protein was suspended overnight with gentle shaking at 4 °C. A second round of ITC was repeated beginning with centrifugation at 4 °C, 30,000 x *g* for 15 min ('cold-spin'). Additional cycles of hot- and cold-spins were repeated until no insoluble impurities were observed following centrifugation of the protein solution at 4 °C and the protein pellet following centrifugation at 37 °C was white. Following the final hot-spin, the purified protein pellets were suspended in ice-cold (0 °C), sterile lysis buffer (~15 mL) without addition of protease inhibitor cocktail. The protein solution was incubated at 4 °C with shaking overnight to completely solubilize the polypeptides.

The protein solutions were dialyzed at 4 °C using the SnakeSkin® Pleated Dialysis Tubing with a molecular weight cut-off of 10 kDa (Pierce Protein Research Products, Thermo Fisher Scientific Inc., Rockford, IL) against distilled, deionized water (5 x 4 L) to remove excess salts. The protein solutions were further purified via sterilization at 4 °C through a 5.0 μM Durapore® membrane filter (Millipore; Billerica, MA) using a Fisherbrand® glass microanalysis vacuum filter holder assembly. Lyophilization of the frozen dialysate produced the elastin-mimetic polypeptides as white spongy solids.

Differential scanning calorimetry

The inverse temperature transitions of the purified amphiphilic block copolymers were monitored using an ultra-sensitive differential scanning calorimeter (VP-DSC from MicroCal,

LLC; Northampton, MA). Dilute solutions of copolymer samples (0.75 to 1 mg/mL) were prepared at 4 °C in sterile ddH_2O and 40 mM NaOH or acetic acid. Samples were degassed under dynamic vacuum 4 °C for 30 min. The thermal transition data was recorded over a temperature range from 4 °C to 60 °C at a scan rate of 1 °C/min with a pre-scan thermostat of 15 min. Reversibility was investigated by cooling and re-scan of samples *in situ*. Data was processed and analyzed using the ORIGIN software (MicroCal, LLC; Northampton, MA).

Results and Discussion

Synthetic genes were designed for the production of elastin-mimetic block copolymers. The genes were based on the pentapeptide repeat of elastin (Val/Ile-Pro-Xaa-Yaa-Gly) in which the phase behavior and mechanical properties depend critically on the identity of the residues within the repeat sequence[29]. Namely, substitution at the fourth residue (Yaa) modulates inverse temperature transition T_t of the aqueous polymer solutions in a manner commensurate with the effect of the polarity of the amino acid side chain on the polymer-solvent interaction. In addition, the presence of an alanine (Ala) residue in place of the consensus glycine (Gly) at the third (Xaa) position of the repeat results in a change in the mechanical response of the material from elastomeric to plastic. The macroscopic effects of these alterations were employed to design amphiphilic elastin-mimetic polypeptide sequences, which emulate **BAB**-triblock copolymers that undergo microscopic phase separation to afford thermoplastic elastomer gels.

The seamless cloning technique described here was used to successfully generate large, full-length DNA concatemers separately encoding the hydrophilic, elastic (**S1**) and the hydrophobic, plastic (**S3)** domains (Scheme 1). The process utilizes the recognition/cleavage patterns of the type IIs restriction endonucleases, in which the cleavage of the DNA duplex occurs at a specific position that is downstream of its recognition site. Specifically, we employed endonucleases that have at least a six-nucleotide recognition sequence and cleave downstream at a defined position to generate 5' cohesive-ended fragments with single-stranded overhangs of four bases. Examples of commercially available type IIs endonucleases that exhibit this restriction pattern are shown in Table 1. For this work, we have utilized the enzymes *Bbs* I and *Bsm*B I. These endonucleases were selected for their commercial availability, differences in recognition site, and compatibility with the cloning and expression plasmids.

Table 1. Recognition and cleavage patterns of commercially available type IIs restriction endonucleases. The enzymes employed for these studies are highlighted.

Enzyme	Recognition sequence	Cleavage pattern
*Aar*I	5'-CACCTGC-3'	(+4/+8)
*Acc*36I	5-ACCTGC-3'	(+4/+8)
*Ace*III	5'-CAGCTC-3'	(+7/+11)
*Bau*I	5'-CACGAG-3'	(–5/–1)
*Bbs*I	5'-GAAGAC-3'	(+2/+6)
*Bfu*AI	5-ACCTGC-3'	(+4/+8)
*Bsa*I	5'-GGTCTC-3'	(+1/+5)
*Bse*Y1	5'-CCCAGC-3'	(–5/–1)
*Bsm*BI	5'-CGTCTC-3'	(+1/+5)
*Bsp*MI	5-ACCTGC-3'	(+4/+8)
*Bte*ZI	5'-GCGATG-3'	(+10/+14)
*Eco*31I	5'-GGTCTC-3'	(+1/+5)

Annealing of two pairs of synthetic oligonucleotides together over a period of 3 hours assembled the duplex DNA monomers, **S1** and **S3** (Scheme 1). The presence of the *Hin*d III and *Bam*H I restriction sites at the 5' and 3' termini, respectively, facilitate insertion of the DNA monomers into the multiple cloning region of the plasmid pZErO®-2.1 (Figure 3). The type IIs restriction endonuclease sites for *Bbs* I and *Bsm*B I were placed near the termini of the DNA cassette at internal positions, which mark the beginning and end, respectively, of the coding sequence of the peptide repeat motif. The elastin monomers were isolated from the cloning plasmid via sequential digestion with *Bbs* I and *Bsm*B I. Cleavage with either enzyme produces complementary, non-palindromic cohesive ends comprising the four-base sequence, 5'-TCCA-3' in the forward (sense) direction and 5'-TGGA-3' in the reverse (anti-sense) direction (Scheme 1). The digested DNA monomers are capable of self-ligation in a head-to-tail fashion to generate a library of concatemers (Figure 4[9]) that differ by units of 75 base pairs corresponding to the length of the DNA monomer coding sequence.

The cloning plasmids **pSCP1** and **pSCP2**, which contain the DNA monomer, retain the restriction sites for *Bbs* I and *Bsm*B I. Digestion of these acceptor plasmids by either endonuclease results in cohesive ends that are compatible with directional insertion of the DNA concatemers. Following electrophoretic size fractionation, the pool of concatemers of desired size range is subcloned into the singly cut cloning plasmids. The recombinant clones are screened for concatemer inserts of desired size (Figure 5[9]) and propagated for use in subsequent construction of the block cassettes. Two recombinant clones each were selected following the **S1** and **S3** DNA cassette concatemerization procedure and propagated for use in subsequent cloning steps (Figure 6). Using this procedure, concatemers can regularly be selected that are 1000 to 3000 base pairs in length. In addition, the recombinant concatemer plasmids **pSCP3** and **pSCP4**

Comments for pZErO™-2.1
3297 nucleotides

Lac Promoter/Operator Region: bases 95-216
M13 Reverse Priming Site: bases 205-221
LacZα ORF: bases 217-558
Sp6 Priming Site: bases 239-256
Multiple Cloning Site: bases 269-381
M13 (-20) Forward Priming Site: bases 415-430
M13 (-40) Forward Priming Site: bases 434-450
Fusion Joint: bases 559-567
*ccd*B Lethal Gene ORF: bases 568-870
f1 origin: bases 895-1307
Kanamycin Resistance ORF: bases 2116-1322
ColE1 origin: bases 2502-3175

Invitrogen life technologies

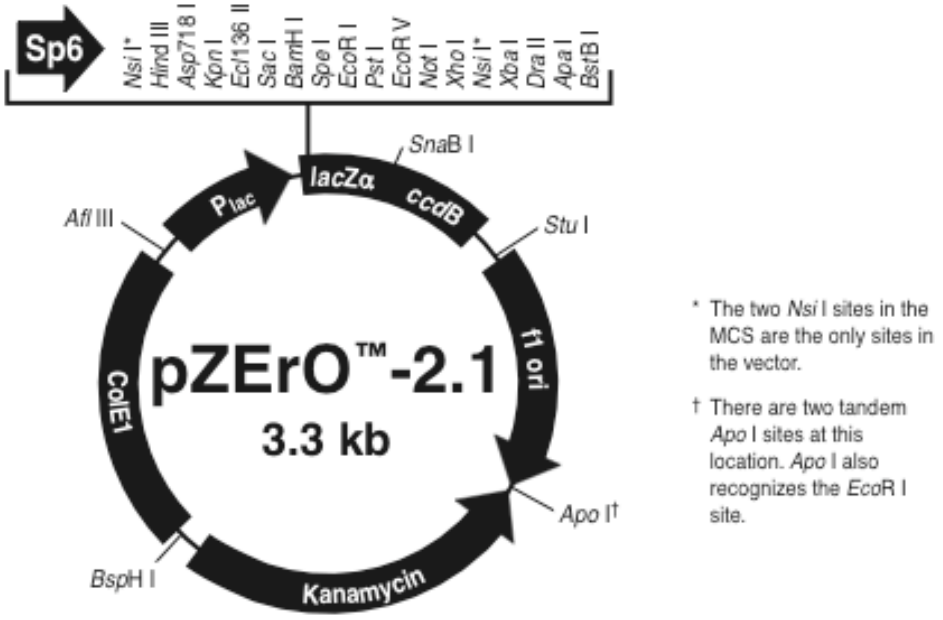


Figure 3. Plasmid map of the commercial cloning vector pZErO®-2.1. The restriction endonuclease recognition sites within the multiple cloning site are shown in detail.

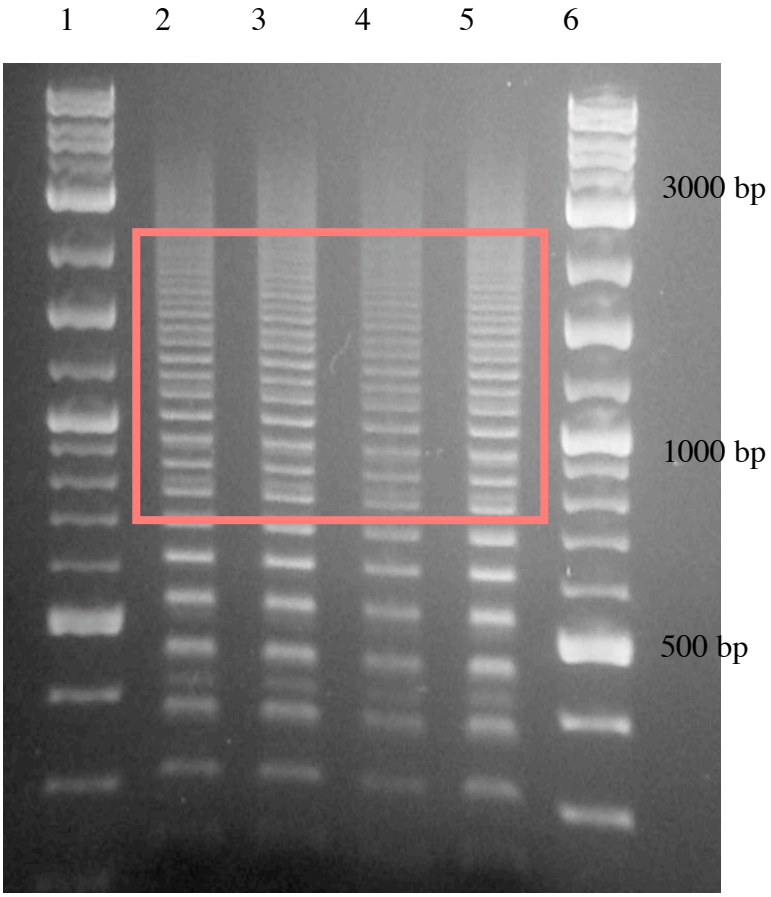


Figure 4. A 1% agarose gel showing the results of multimerization of the elastin **S1** monomer DNA. Lane 1: 1 kb and 100 bp (1 μL each) DNA ladder from NEB (Beverly, MA); lanes 2-5: Multimerization reaction; and lane 6: 1 kb and 100 bp (1 μL each) DNA ladder. The red box is drawn around the region of the gel that was excised for isolation of multimers in the range of 1000 to 3000 bp in length.

A similar ladder of concatemerization products is observed for ligation of **S3** monomers (*data not shown*).

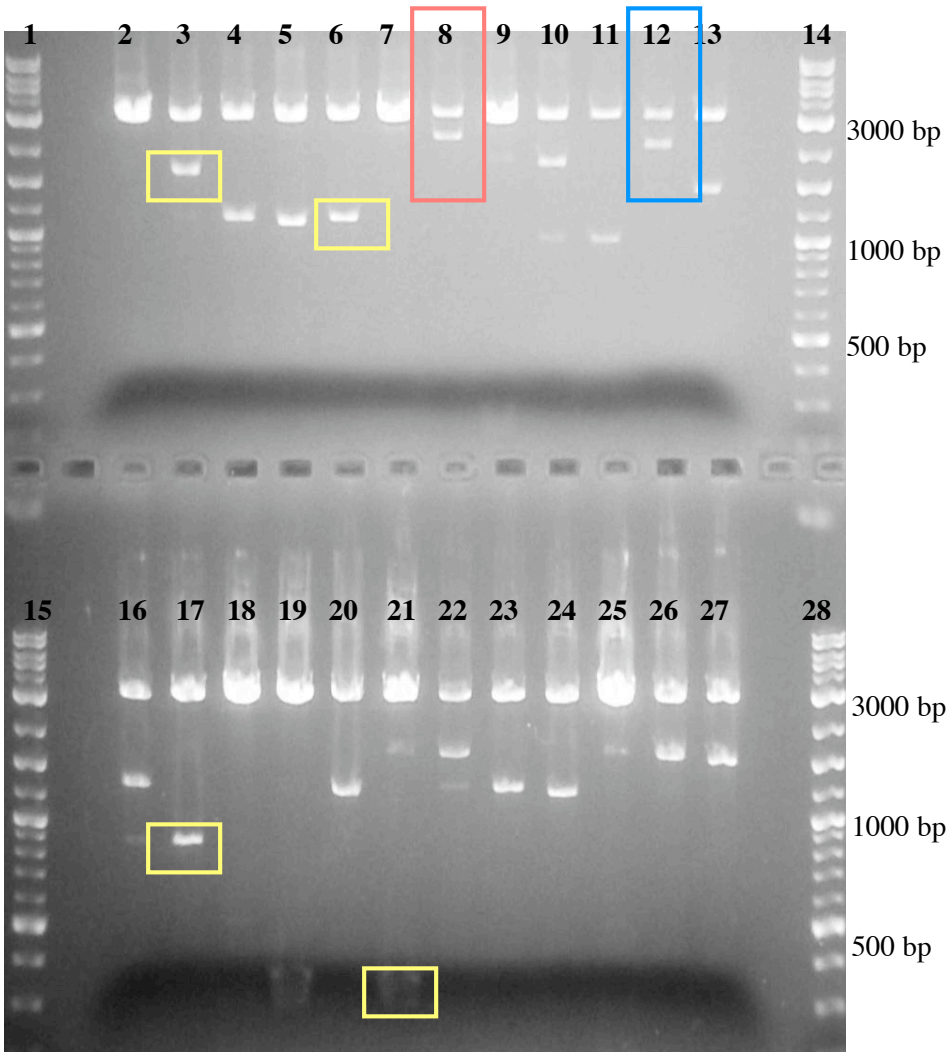


Figure 5. A 1 % agarose gel demonstrating the results of multimerization of the **S3** gene. DNA was isolated from several separate colonies grown on selective plates after transformation with plasmids containing **S3** multimers of various lengths. The DNA (~500 ng) was digested with *Bam*H I and *Hind* III in 24 separate reactions and 20 μL of each reaction were run on the gel. Several multimers of variable lengths were chosen for cloning and future experiments. The red and blue boxed regions correspond to the DNA concatemer clones **pSCP4.7** and **pSCP4.11**, respectively, that were isolated for subsequent cloning and experiments,. The yellow boxed regions correspond to multimers of the following approximate sizes, 1800 bp, 1200 bp, 900 bp, and 350 bp. Lanes 1,14, 15, 28: 1 kb and 100 bp (1 μL each) DNA ladder (NEB); and lanes 2-15, 16-27: The DNA from 24 single colonies, each containing the pZErO®-2 plasmid and **S3** multimer insert.

A similar distribution of recombinant DNA concatemer clones is observed for multimerization of the **S1** gene (*data not shown*).

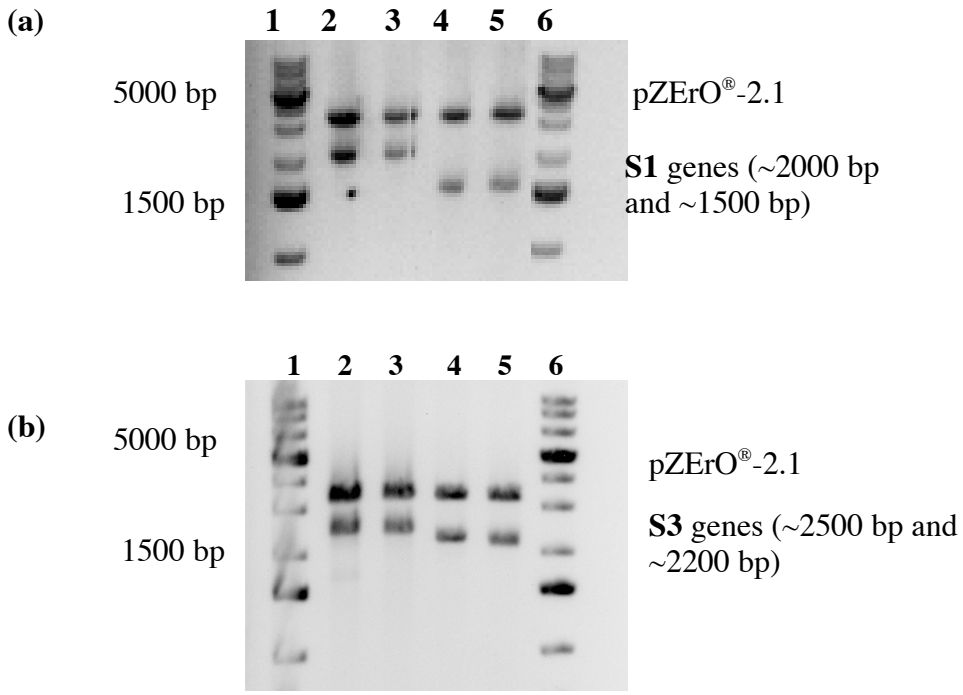


Figure 6. (a) A 1 % agarose gel depicting plasmid **pSCP3.1** and **pSCP3.11** double digested (~500 ng plasmid DNA) with *Bam*H I and *Hin*d III indicating the **S1** multimer gene inserts (~2000 and 1500 bp). Twenty microliters of each digestion reaction were run on the gel in separate lanes. Lane 1,6: O'GeneRuler™ 1 kb Plus DNA ladder (Fermentas Inc., Glen Burnie, MD); lanes 2,3: The DNA from 2 single colonies both containing **pSCP3.1**; and lanes 4,5: The DNA from 2 single colonies both containing **pSCP3.11**. (b) 1 % agarose gel depicting plasmid **pSCP4.7** and **pSCP4.11** double digested with *Bam* HI and *Hin*d III indicating the **S3** multimer gene inserts (~2500 and 2200 bp). Lane 1,6: O'GeneRuler™ 1 kb Plus DNA ladder; lanes 2,3: The DNA from 2 colonies both containing **pSCP4.7**; and lanes 4,5: The DNA from 2 colonies both containing **pSCP4.11**.

retain the *Bbs* I and *Bsm*B I restriction sites at the beginning and end, respectively, of the coding sequence of the polypeptide. The cohesive ends of the two different concatemers of **S1** and **S3** are identical and therefore compatible for ligation to each other.

The seamless cloning strategy was also successfully employed for the selective ligation of the pair of concatemers to create synthetic DNA fusions that encode blocks of sequence-repetitive polypeptides as shown in Figures 7 and 8. The plasmids utilized and constructed in the synthesis of the triblock polypeptides are described in Table 2. For the synthesis of artificial genes encoding **BAB**-type triblock polypeptides, two independent ligation reactions were performed to first generate an **AB**- and **BA**-diblock. The gene assembly involves not only selective use of *Bbs* I and *Bsm*B I, but also the participation of a third enzyme *Nco* I, which cleaves at a position within the kanamycin resistance gene of the plasmid pZErO®-2.1. Each plasmid clone is subjected to restriction digestion with *Nco* I and one of the type IIs restriction endonucleases; *Bsm*B I for the proximal clone and *Bbs* I for the distal clone with respect to the anticipated N terminus within the coding sequence of the polymer product. The digestion reactions cleave each plasmid into two pieces that can be separated via agarose gel electrophoresis to isolate the concatemer fragments. Ligation of the fragments not only reconstitutes the kanamycin resistance marker, but also generates a direct fusion of the proximal and distal concatemer, affording constructs **pSCP5** and **pSCP6** that encode the **AB**- and **BA**-diblock polypeptide, respectively. The cloning method proceeded with very high efficiency and nearly eliminated the occurrence of false positives due to antibiotic selection of recombinant clones with reconstitution of the selectable marker. The diblock cassette is flanked by unique *Bbs* I and *Bsm*B I sites that are employed for further rounds of directional cloning to generate the **BAB**-type triblock copolymers.

Table 2. List of recombinant plasmids utilized and constructed in the synthesis of the amphiphilic triblock (**BAB**-type) copolymers, **B10** and **B11**.

Recombinant Plasmid Name	**Plasmid/Insert**	**Comments**
pSCP1	pZErO®-2.1/elastic (**S1**) DNA monomer	Zeo^R; Reference[9]
pSCP2	pZErO®-2.1/plastic (**S3**) DNA monomer	
pSCP3	pZErO®-2.1/elastic (**S1**) concatemer	Two variants of ~2000 bp and 1500 bp
pSCP4	pZErO®-2.1/plastic (**S3**) concatemer	Two variants of ~2500 bp and 2200 bp
pSCP5	pZErO®-2.1/**S1**-**S3** diblock gene fusion	
pSCP6	pZErO®-2.1/**S3**-**S1** diblock gene fusion	
pSCP7	pBAD/His A/**Elastin Adaptor**	Amp^R
pMAP3	pZErO®-2.1/**B10**	Triblock gene fusion
pMAP5	pSCP7/**B10**	
pMAP6	pET-24d (+)/**B10**	Kn^R; Amphiphilic triblock (**BAB**-type) expression plasmid;
pMAP10	pZErO®-2.1/**B11**	Triblock gene fusion
pMAP11	pSCP7/**B11**	
pMAP12	pET-24d (+)/**B11**	Amphiphilic triblock (**BAB**-type) expression plasmid

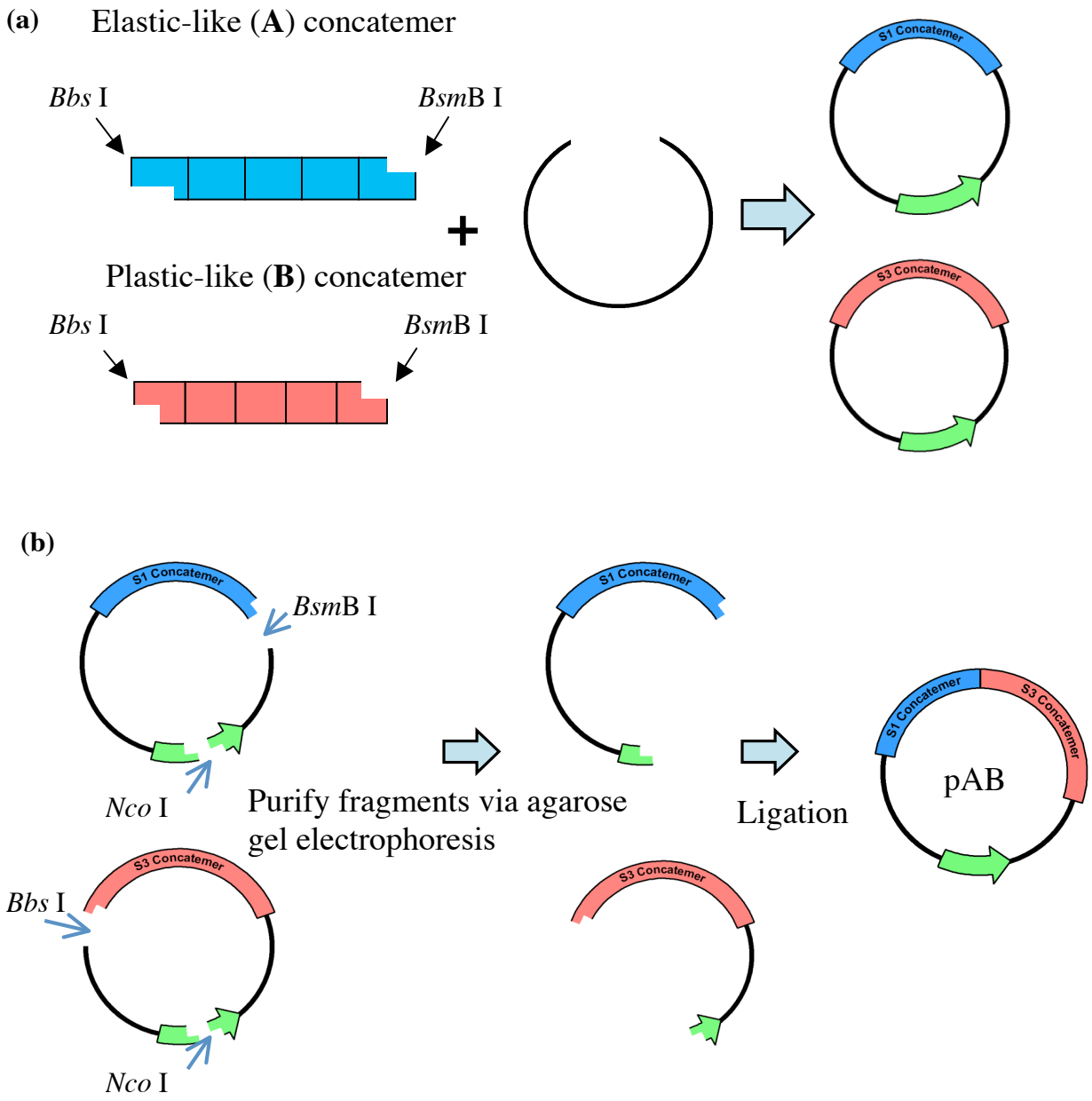


Figure 7. Directional cloning strategy for assembly of synthetic genes encoding block polypeptides of elastin-mimetic sequence repeats. (a) Plasmid clones encoding the elastic (**S1**) and plastic (**S3**) concatemers are cloned into the pZErO®-2.1 plasmid. (b) The DNA concatemer plasmids are then assembled to create an **AB**-diblock and **BA**-diblock gene fusion.

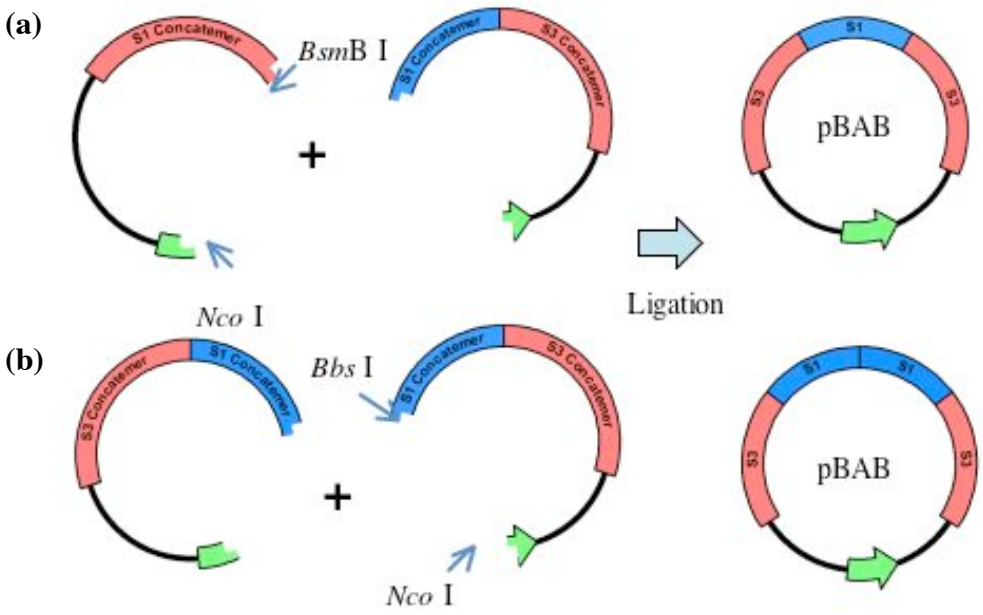


Figure 8. The plasmids containing the **AB** and **BA**-diblock gene fusions were used as starting material for the directional cloning of two triblock **BAB**-type gene fusions, which encode amphiphilic copolymer (a) **B10** and (b) **B11**. The amphiphilic triblock copolymer **B10** comprises an elastic midblock flanked on either side by the plastic endblocks, while **B11** contains an elastic midblock twice as large.

The diblock plasmids **pSCP5** and **pSCP6** were propagated and utilized as starting material for construction of the triblock gene fusion along with the plastic concatemer containing plasmid **pSCP4**. In a second round of directional cloning, two separate ligation reactions were performed where the **AB**-diblock construct represents the distal block and the **S3**-type concatemer or **BA**-diblock represents the proximal blocks (Figure 8). Successful cloning of the triblock gene fusions was identified via agarose gel electrophoresis (Figure 9). Assembly of the triblock gene fusions is completed within the cloning plasmid, generating plasmids **pMAP3** and **pMAP10** (Figure 10), which encode the corresponding amphiphilic block copolymers **B10** and **B11**. The final expression cassette is liberated by sequential restriction digestion with *Bbs* I and *Bsm*B I. The concatemeric DNA cassette is then cloned into an appropriately modified expression vector. Since the non-palindromic cohesive ends of the concatemeric cassettes are not compatible with the cohesive ends generated by restriction digestion at the endonuclease recognition sites typically found in the multiple cloning sites of commercially available expression plasmids, the expression vector is modified with an adaptor sequence. The pBAD/His A expression plasmid (Figure 11) was modified through the insertion of a synthetic adaptor sequence (Scheme 2) between the *Nco* I and *Hin*d III sites of the polylinker. The synthetic adaptor sequence has two *Bbs* I restriction sites. Therefore digestion with the type IIs endonuclease generates 5-cohesive ends compatible with those of the DNA concatemers. It is important to note that the restriction sites are internally oriented with respect to the polylinker, and therefore digestion at both sites removes a small DNA fragment containing the cleavage site. In addition, a recognition/cleavage site for *Kpn* I is placed in between the two *Bbs* I sites. Digestion with *Kpn* I prevents unwanted side reactions, such as re-ligation of the polylinker plasmid in cases where only one *Bbs* I site was cleaved. Treatment of the modified pBAD/His A

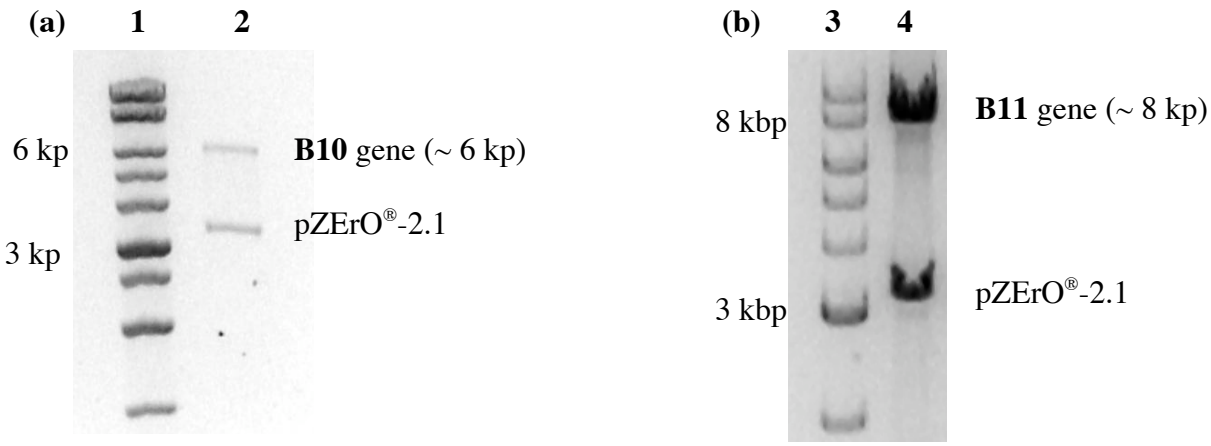


Figure 9. Two 0.75 % SeaPlaque® GTG® agarose gels depicting plasmids (a) **pMAP3** and (b) **pMAP10** double digested with *Bam*H I and *Hin*d III indicating the **BAB**-triblock gene inserts, which encode the amphiphilic copolymers **B10** and **B11**, respectively. Twenty microliters of each digestion reaction were run on two separate gels. (a) Lane 1: 1 kb and 100 bp (1 μL each) DNA ladder (NEB) and lane 2: The DNA from a single colony containing the pZErO®-2.1 plasmid and **B10** triblock gene insert. (b) Lane 3: 1 kb and 100 bp (1 μL each) DNA ladder (NEB) and lane 4: The DNA from a single colony containing the pZErO®-2.1 plasmid and **B11** triblock gene insert.

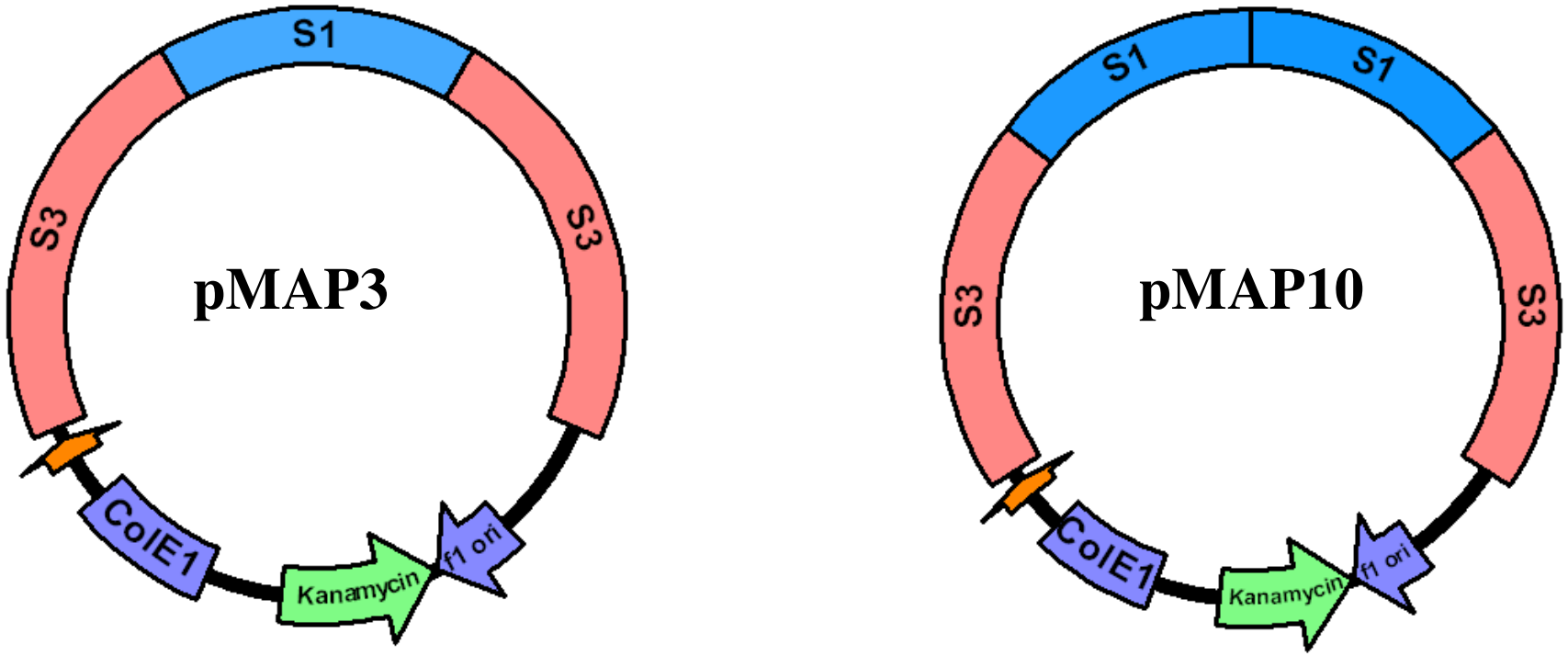


Figure 10. Plasmids **pMAP3** and **pMAP10** represent the recombinant pZErO®-2.1 plasmids containing triblock (**BAB**-type) gene fusions derived from elastin-mimetic sequences, which encode amphiphilic copolymers **B10** and **B11**, respectively.

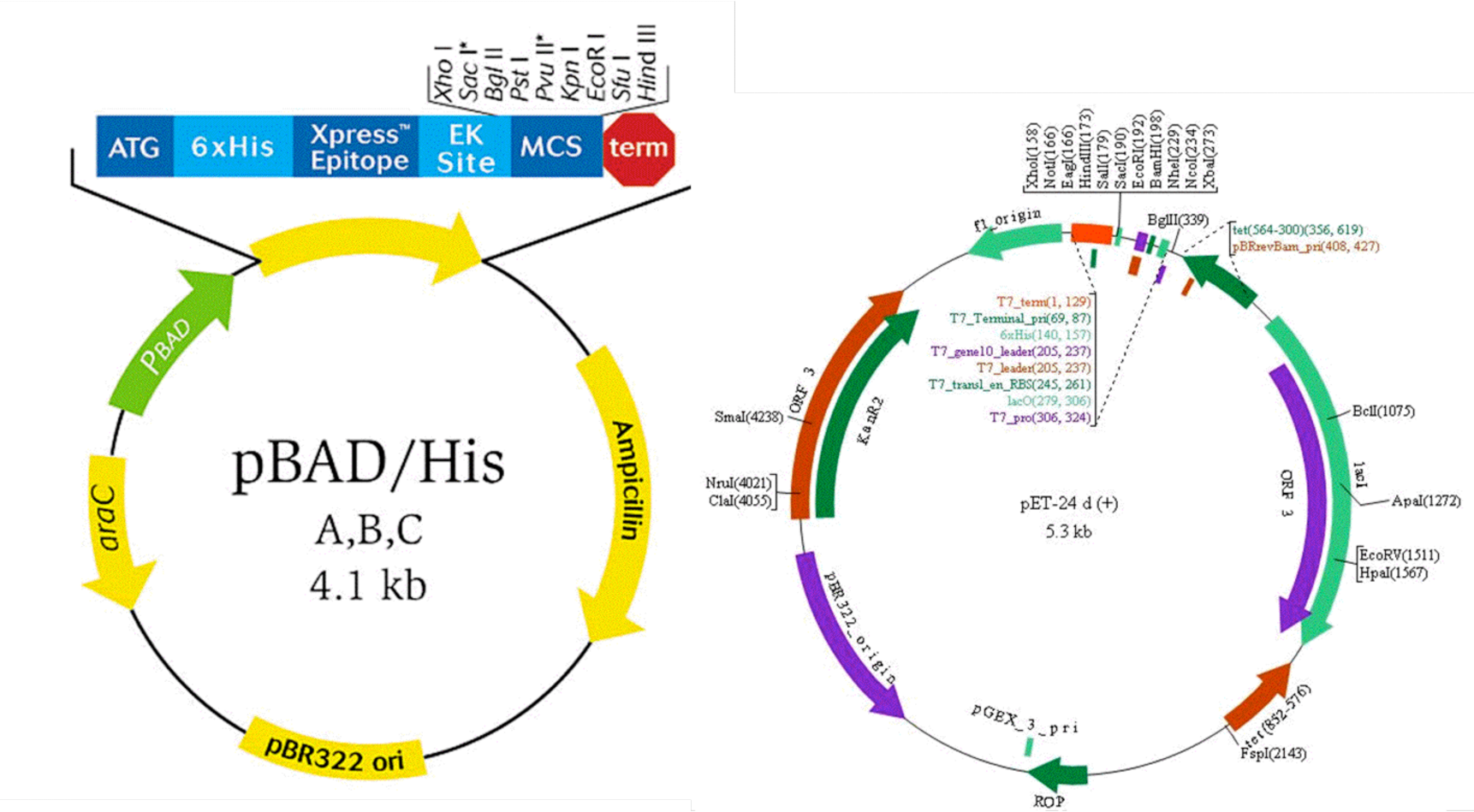


Figure 11. Plasmid maps of the commercially available pBAD/His A and pET-24 d (+) expression vectors utilized for cloning the triblock copolymer expression plasmids. The restriction endonuclease recognition sites within the multiple cloning site of each plasmid are shown in detail.

plasmid **pSCP7** with both enzymes should reduce background levels of undesired cloning products, which may only be discovered through DNA sequencing. Lastly, the sequence of the adapter is configured such that the concatemer cassettes would be presented in the appropriate reading frame for production of the desired triblock polypeptides upon gene expression.

The pET-series of expression plasmid has been utilized to direct overexpression of highly repetitive, heterologous polypeptides in *E. coli* strains such as BL21(DE3) and BLR(DE3)[30]. To take advantage of this protocol, the DNA concatemers were removed from the modified pBAD/His A plasmid as a cassette flanked by the **Elastin Adaptor** sequence and inserted into the multiple cloning site of the pET-24 d (+) expression plasmid between the *Nco* I and *Hin*d III sites. The **B10** and **B11** triblock expression plasmids were designated **pMAP6** and **pMAP12**, respectively (Figure 12). Immediately following this study, a new adapter sequence **Elastin Adaptor2** was generated for direct cloning of DNA cassettes like those described here into pET-24 d (+). The sense (5'-CATGGTTCCAAGAGACCAGGTACCGGTCTCGTCCAGGTGTACG CTAATA-3') and anti-sense (5'-AGCTTATTAGCCTACACCTGGACGAGACCGGTACCTG GTCTCTTGGAAC-3') strands of the adapter were annealed and ligated into the *Nco* I and *Hin*d III sites of the pET-24 d (+) multiple cloning site. Cloning of the DNA cassettes requires digestion of the **Elastin Adaptor2** with the type IIs restriction endonucleases *Bsa* I and *Kpn* I. The enzyme *Bsa* I was chosen for this procedure since both *Bbs* I and *Bsm*B I cleave at one or more sites along the pET-24 d (+) vector DNA. Insertion of the **Elastin Adapter2** into the pET-24 d (+) plasmid will eliminate the intermediate pBAD/His A cloning step described above and prove useful for future cloning of the DNA concatemer cassettes. The modified pET-24 d (+) plasmid was designated **pMAP4**.

Plasmids **pMAP6** and **pMAP12** were transformed into the *E. coli* strain BL21-Gold

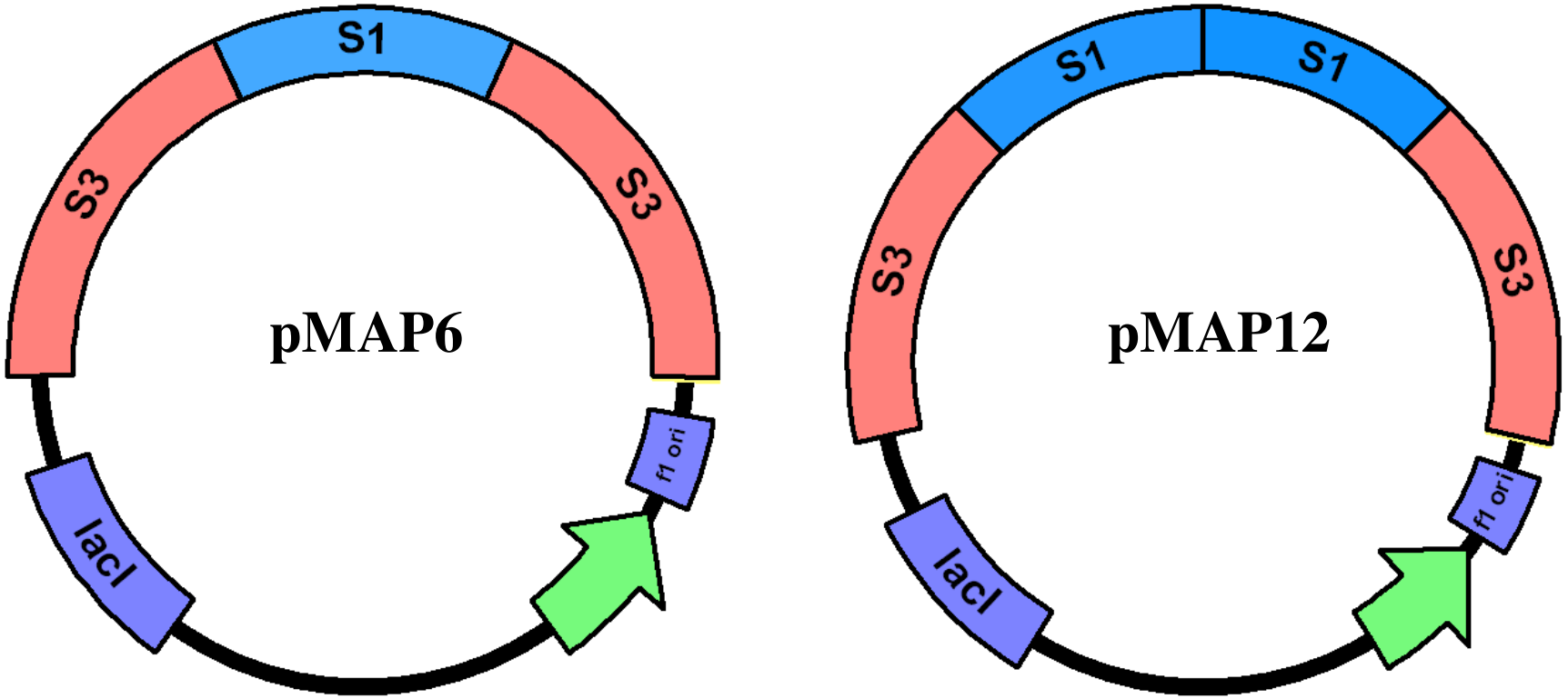


Figure 12. Plasmids **pMAP6** and **pMAP12** represent subcloning of the **B10** and **B11** gene fusions as cassettes flanked by the **Elastin Adaptor** (shown in yellow) into the expression plasmid pET-24 d (+). A kanamycin antibiotic resistance gene (green) is present for plasmid maintenance during transformation and cell culturing.

(DE3). The cells were incubated at 37 °C for 36 h following a hyperexpression protocol described by Daniell, et al[22]. The amphiphilic triblock copolymers **B10** and **B11** derived from elastin-mimetic sequences could be purified to homogeneity from the endogenous proteins of the bacterial expression host via repetitive cycling through the phase transition. The target proteins undergo selective precipitation above the T_t of the polypeptides at high salt concentration (2 M NaCl)[31]. For isolation of **B10** and **B11** with high purity, three to five cycles were necessary. Dialysis, cold filter sterilization, and lyophilization afforded proteins **B10** and **B11** as fibrous solids in high yields (250 and 260 mg/L, respectively).

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS PAGE) analysis was used to detect the accumulation of the target protein fusions during the hyperexpression protocol (Figure 13). Whole cell lysates of the elastin-mimetic polymer expression cultures were separated using a 7% gel and the target protein bands were visualized by negative staining with zinc sulfate (Figure 13). As previously reported, the molar masses of the amphiphilic block copolymers as observed by SDS PAGE appear to be ~20% larger than the calculated molar masses[32,33]. The amino acid sequences of the two elastin-mimetic triblock copolymers are represented in Scheme 3. Collaborative studies with the Chaikof lab confirmed the identity of copolymer **B10** by MALDI-TOF mass spectrometry and amino acid compositional analysis[34]. Current efforts are underway to characterize the primary structure of **B11**.

Differential scanning calorimetry (DSC) on dilute aqueous solutions (approximately 1 mg/mL) of the **B10** and **B11** copolymers displayed sharp endothermic transitions at approximately 21 °C (Figure 14), which were reversible as demonstrated by cooling and rescan *in situ*. From the DSC results, the calculated van't Hoff enthalpies for **B10** (+480.5 ± 81.9 kJ mol^{-1}) and **B11** (+639.1 ± 19.5 kJ mol^{-1}) copolymers are consistent with previously observed

endothermic transition enthalpies (ΔH) for elastin-mimetic polymers[29] and block copolymers[35]. In addition, DSC data for both triblock copolymers at high and low pH revealed endothermic thermal transitions also between 19 °C and 20 °C (Figures 15 and 16). The lack of significant pH dependence on the thermal transition temperature of the polypeptides suggests that the glutamic acid residues in the **S1** central blocks of these copolymers are not involved in the transition. The consistency of the phase transitions of both triblock polypeptides suggests that this process involves selective desolvation and micro-phase separation of the endblocks.

Scheme 3. Amino Acid Sequence of Protein-Based Block Copolymers **B10** (n = 27) and **B11** (n = 45)

B10/**B11** = {[VPAVG(IPAVG)$_4$[(IPAVG)$_5$]$_{28}$-[X]-{[VPAVG(IPAVG)$_4$][(IPAVG)$_5$]$_{26}$IPGVG,

where [X] = (IPAVG)(VPGAG)VPGEG(VPGAG)$_2$[(VPGAG)$_2$VPGEG(VPGAG)$_2$]$_n$

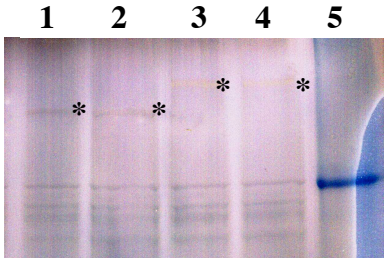


Figure 13. SDS PAGE analysis of triblock (**BAB**-type) copolymers run on a 7% gel and negatively stained with zinc sulfate. The expected molecular weights of **B10** and **B11** are approximately 178 kDa and 214 kDa, respectively. Lanes 1,2: whole cell lysates from **B10** hyperexpression at 24 and 36 hr, respectively; lanes 3,4: whole cell lysates from **B11** hyperexpression at 24 and 36 hr, respectively; and lane 5: 8 μL of the Prestained Protein Marker, Broad Range (10-175 kDa) from NEB (Beverly, MA). Protein bands are indicated by asterisks (*).

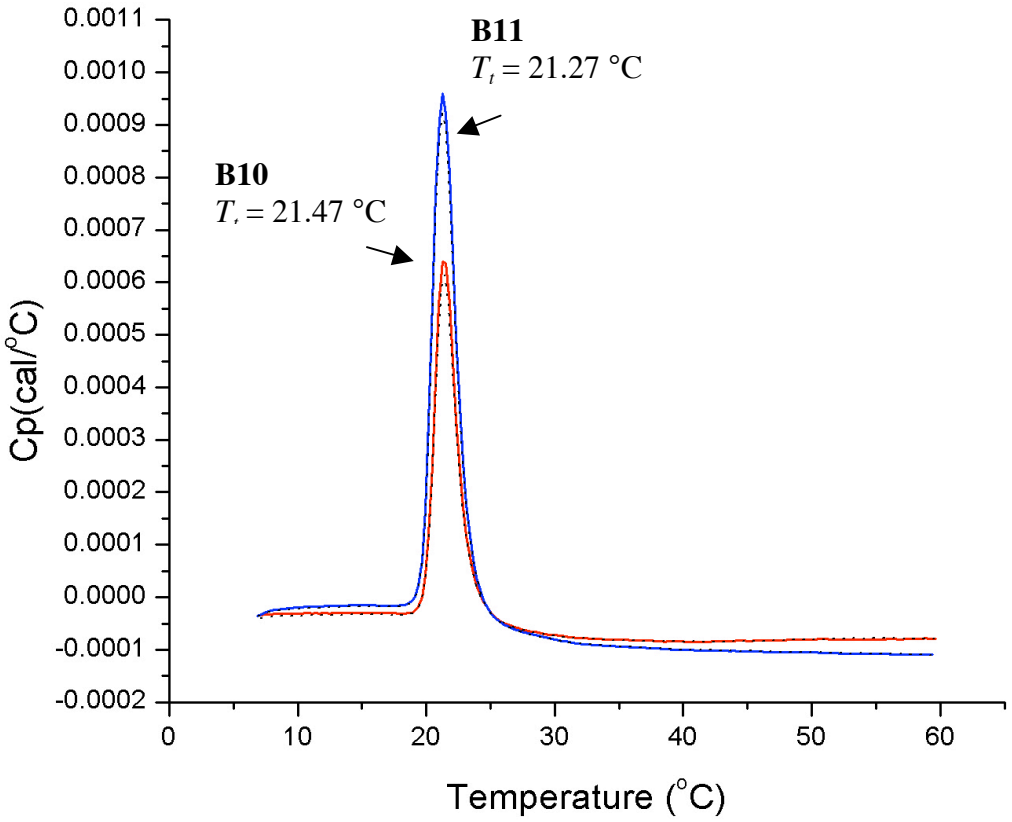


Figure 14. Raw DSC data for the endothermic thermal transitions of elastin-mimetic triblock copolymers **B10** (red) and **B11** (blue) in dilute aqueous solutions (~1 mg/mL). The reversibility of the transition is demonstrated by rescan of the samples (shown by dotted lines, which overlap with curves for initial thermal transitions).

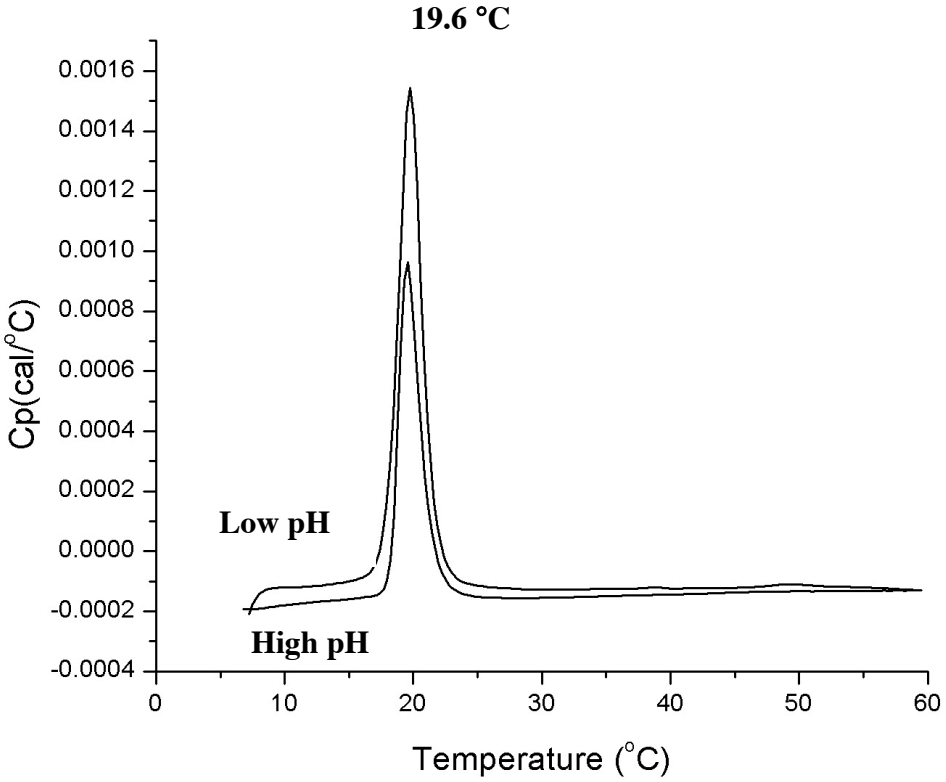


Figure 15. Raw DSC data for the endothermic thermal transitions of **B10** in dilute aqueous solutions (1 mg/mL) at low pH (40 mM acetic acid) and high pH (40 mM NaOH). The temperature maximum for the thermal transition under both acidic and basic buffer conditions corresponds to 19.6 °C.

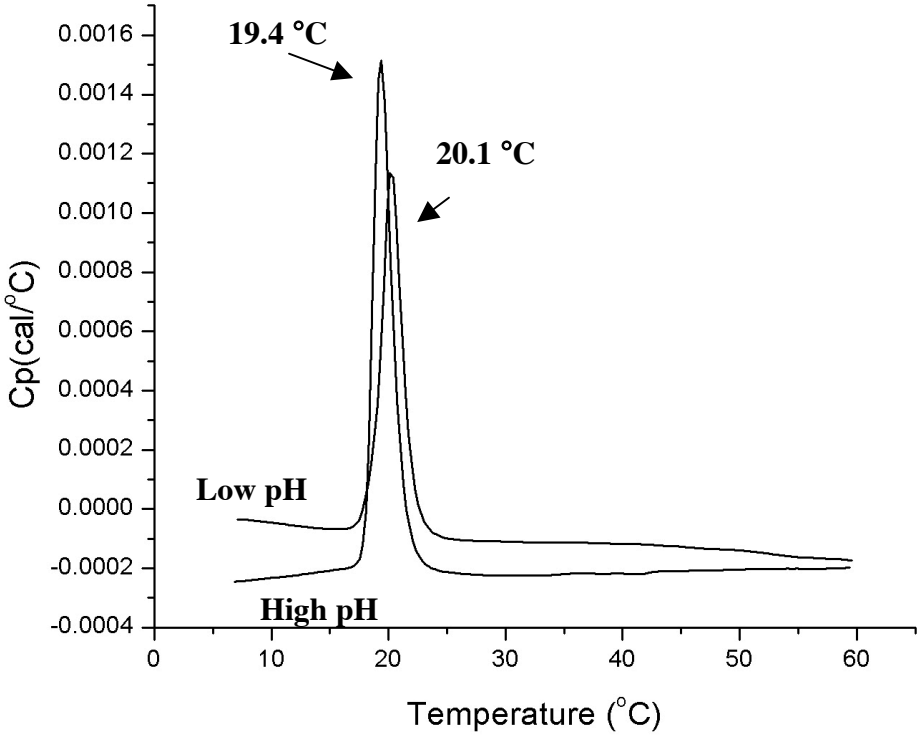


Figure 16. Raw DSC data for the endothermic thermal transition of **B11** in dilute aqueous solutions (1 mg/mL) at low pH (40 mM acetic acid) and high pH (40 mM NaOH). The temperature maximum for the transition under each pH condition is indicated.

Conclusion

The cloning strategy described allowed for synthesis of (**BAB**-type) triblock copolymers derived from elastin-mimetic polypeptide sequences in a manner that is modular and convergent. The seamless cloning strategy permitted the directional assembly of the DNA concatemer products in construction of genetic fusions up to 8000 base pairs in length. Using the hyperexpression protocol, we were able to generate high yields of the recombinant polypeptides and easily purify the elastin-mimetic polymers from the endogenous host proteins to homogeneity. SDS-PAGE analysis provided evidence of the expression of the full-length proteins. The triblock elastin-mimetic polypeptides exhibited an endothermic thermal transition characteristic of native elastin and other elastin-mimetic polypeptides. The temperature of the inverse phase transition occurs at a T_t below room temperature, which reflects a temperature-dependent assembly behavior predicted for the hydrophobic plastic domains. Furthermore, the temperature at which the inverse phase transition occurs is independent of the pH of the copolymer solution. These results indicate a process consistent with hydrophobic or other solvation-related interactions. The cloning strategy should be applicable for the construction of artificial proteins comprising a wide variety of sequence-repetitive polypeptides, including those of highly complex macromolecular architecture. The seamless cloning strategy and expression protocol can be easily applied for the design and synthesis of multi-block copolymers composed of an even more diverse set of blocks. Blocks encoding repetitive polymers derived from other native-protein based materials, such as silk and other extracellular proteins like collagen and keratin, as well as bioactive polypeptide sequences, which may promote cell adhesion and migration, can be joined to generate multi-domain polymers, where each domain displays unique structures, mechanical properties, and biological function.

References

[1] Goeden-Wood, N. L.; Conticello, V. P.; Muller, S. J.; Keasling, J. D. *Biomacromolecules* **2002**, *3*, 874-879.

[2] Tirrell, D. A.; Fournier, M. J.; Mason, T. L. *MRS Bulletin* **1991**, *16*, 23-28.

[3] Van Hest, J. C. M.; Tirrell, D. A. *Chemical Communications* **2001**, *4*, 1897-1904.

[4] Tirrell, J. G.; Tirrell, D. A.; Fournier, M. J.; Mason, T. L. In *Protein-based materials*; McGrath, K. P., Kaplan, D. L., Eds.; Birkhäuser: Boston, 1997, p 61-99.

[5] Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Biotechnology Progress* **1990**, *6*, 198-202.

[6] Urry, D. W.; Gowda, D. C.; Parker, T. M.; Luan, C.-H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D. *Biopolymers* **1992**, *32*, 1243-1250.

[7] Urry, D. W.; Luan, C.-H.; Parker, T. M.; Gowda, D. C.; Prasad, K. U.; Reid, M. C.; Safavy, A. *Journal of the American Chemical Society* **1991**, *113*, 4346-4347.

[8] Lixin, M. *Biomacromolecules* **2006**, *7*, 2099-2107.

[9] Payne, S. C.; Patterson, M.; Conticello, V. P. In *Protein Engineering Handbook*; Lutz, S., Bornscheuer, U. T., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2009; Vol. 1, p 915-936.

[10] McMillan, R. A.; Caran, K. L.; Apkarian, R. P.; Conticello, V. P. *Macromolecules* **1999**, *32*, 9067-9070.

[11] McMillan, R. A.; Conticello, V. P. *Macromolecules* **2000**, *33*, 4809-4821.

[12] McMillan, R. A.; Lee, T. A. T.; Conticello, V. P. *Macromolecules* **1999**, *32*, 3643-3648.

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

[14] Wright, E. R.; Conticello, V. P. *Advanced Drug Delivery Reviews* **2002**, *54*, 1057-1073.

[15] Wright, E. R.; Conticello, V. P.; Apkarian, R. P. *Microscopy and Microanalysis* **2003**, *9*, 171-182.

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

[17] Cappello, J.; Ferrari, F. A. In *Plastics from microbes: microbial synthesis of polymers and polymer precursors*; Mobley, D. P., Ed.; Hanser/Gardner Publications: Munich, 1994, p 35-92.

[18] Ferrari, F. A.; Cappello, J. In *Protein-based materials*; McGrath, K. P., Kaplan, D. L., Eds.; Birkhäuser: Boston, 1997, p 37-60.

[19] Heslot, H. *Biochimie* **1998**, *80*, 19-31.

[20] Tirrell, J. G.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Chemistry and Engineering News* **1994**, *72*, 40-51.

[21] Padgett, K. A.; Sorge, J. A. *Gene* **1996**, *168*, 31-35.

[22] Daniell, H.; Guda, C.; McPherson, D. T.; Zhang, X.; Xu, J.; Urry, D. W. *Methods in Molecular Biology* **1997**, *63*, 359-371.

[23] Chilkoti, A.; Christensen, T.; MacKay, J. A. *Current Opinion in Chemical Biology* **2006**, *10*, 652-657.

[24] Lee, T. A. T.; Cooper, A.; Apkarian, R. P.; Conticello, V. P. *Advanced Materials* **2000**, *12*, 1105-1110.

[25] Zhou, Y.; Wu, S.; Conticello, V. P. *Biomacromolecules* **2001**, *2*, 111-125.

[26] Sambrook, J.; Russell, D. W. *Molecular cloning: a laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

[27] Simpson, R. J. In *Proteins and Proteomics*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2003.

[28] Meyer, D. E.; Chilkoti, A. *Nature Biotechnology* **1999**, *17*, 1112-1115.

[29] Urry, D. W.; Luan, C.-H.; Harris, C. M.; Parker, T. M. In *Protein-based materials*; McGrath, K. P., Kaplan, D. L., Eds.; Birkhäuser: Boston, 1997, p 133-178.

[30] Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W. *Methods in Enzymology* **1989**, *185*, 60-89.

[31] McPherson, D. T.; Morrow, C.; Minehan, D. S.; Wu, J.; Hunter, E.; Urry, D. W. *Biotechnology Progress* **1992**, *8*, 347-352.

[32] Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2002**, *3*, 357-367.

[33] Trabbic-Carlson, K.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2003**, *4*, 572-580.

[34] Wu, X. Y.; Sallach, R. E.; Caves, J. A.; Conticello, V. P.; Chaikof, E. L. *Biomacromolecules* **2008**, *9*, 1787-1794.

[35] Cooper, A. *Biophysical Chemistry* **2000**, *85*, 25-39.

Chapter 3

Introduction of alkene functionality into elastin-mimetic polypeptides via an *E. coli* expression system

Introduction

The engineering of polymers from conventional organic monomers allows chemists to prepare macromolecules with a wide variety of functional groups. Such methods, however, cannot achieve the sequence-specificity and monodispersity available by biosynthesis of protein-based polymers. On the other hand, the array of chemical functionality available via protein translation is modest and limited by the small library of amino acids specified by the genetic code. Lately research in the *de novo* synthesis of protein-based polymers has been directed at combining the diversity of synthetic polymer chemistry with the sequence control of protein biosynthesis. The introduction of new building blocks displaying unique functional groups with novel patterns of reactivity significantly expands the scope and impact of protein engineering.

One approach for engineering protein-based polymers consisting of monomers with expanded chemical diversity has been targeted at enhancing the ability of the protein biosynthetic machinery to utilize non-natural amino acid analogues beyond the 20 natural amino acids. Toward this end, particular attention has been directed at the aminoacyl-tRNA synthetases (aaRS), which conjugate amino acids to their cognate tRNAs. Expanding the substrate specificity of the translational apparatus presents unique challenges. Specifically, the discriminatory control of the aminoacyl-tRNA synthetases, which ensures the fidelity of translation of genetic information into protein sequence, limits the variety of amino acid structures that can be exploited during the engineering of natural and artificial proteins *in vivo*. Therefore, several methods have been directed at overcoming the strict specificity of the synthetases. For example, chemisynthetic aminoacylation protocols, developed by Hecht and coworkers[1-3] and utilized by others[4-13], eliminate the use of the synthetases all together, but typically require the use of cell-

free translation methods that are characterized by relatively low yields. In addition, the synthetase activities of the cell may be altered through overexpression of a mutant synthetase or heterologous synthetase to generate an organism with an expanded genetic code[14-17]. In some instances, however, it is possible to take advantage of the inherent promiscuity of the wild-type synthetases, which have been shown to activate and charge substrates other than the native, proteinogenic amino acids[18-21]. Of the available methods, the later biosynthetic method is advantageous because it does not require chemical aminoacylation or the use of cell-free translation protocols. The simplicity of the approach and its capacity to provide relatively large quantities of engineered proteins needed for applications in materials science make this our preferred method of choice for analogue incorporation.

Among the myriad of structurally similar analogues of the canonical amino acids that have exhibited translational activity *in vivo*, we have targeted the methionine derivative, (*2S*)-*2*-amino-5-hexenoic acid (L-**Hag**) (Scheme 1). This analogue is characterized by a terminal vinyl group, which may serve as an attractive site for photo-initiated cross-linking or further derivatization via the versatile chemistry of alkenes[22]. Based on previous reports[18,23-25] of the permissiveness of the *E. coli* methionyl-tRNA synthetase (MetRS), Tirrell and coworkers[26,27] investigated the incorporation of a diverse set of translationally active methionine analogues *E. coli*, including the analogue Hag. It is important to note that in their experiments, cells were supplemented with the analogue as racemic mixture of D- and L-isomers. Their studies demonstrated global incorporation of the analogue at eight sites within a target protein, mouse dihydrofolate reductase (mDHFR), with a high degree of fidelity and in relatively high yields. Furthermore, subsequent studies by the authors showed that the use of a modified expression

stain characterized by increased activity of the methionyl-tRNA synthetase (MetRS) resulted in improved protein yield of the Hag containing protein[28]. The observed promiscuity of the MetRS may result not only from structural and electronic similarities of the methionine analogue, but also from unique features of the aminoacyl-tRNA synthetase. First, comparison of the crystal structures of the free[29] and methionine-bound[30] enzyme suggest that the MetRS has an unique conformation flexibility. In addition, the wild-type aaRS is characterized by reduced editing function.

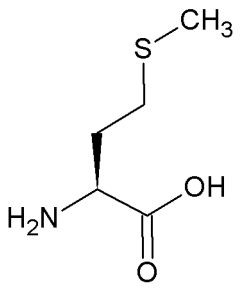
The incorporation of Hag into the polypeptide sequence during *in vivo* protein synthesis in an *E. coli* expression system represents important new opportunities for protein engineering. Incorporation of the alkene functionality introduces a potential site for chemoselective modification to mimic common post-translational modifications[31], olefin cross metathesis[32], or photo-initiated cross-linking. For example, the addition of fatty acids or sugars during *in vivo* protein synthesis is often important for protein function and trafficking. A report by Davis and coworkers described a new glycosylation strategy using thiol-ene coupling between Hag modified proteins and various glycosyl-thiols[31]. In addition, incorporation of Hag is a promising strategy for developing another orthogonal reaction pathway for labeling proteins via olefin metathesis. While metathesis of the unnatural amino acid analogue in organic solvent has been previously reported[33,34], future investigations directed at isolating new water-soluble catalysts could make the strategy an attractive method for cross-linking of Hag-modified proteins or their conjugation to other molecules. Finally, photo-initiated cross-linking of genetically engineered polypeptides containing Hag would facilitate the generation of protein-based biomaterials with improved mechanical and biological properties via incorporation of covalent cross-links. Recent

developments in biomaterials have realized the strong effect of mechanical properties on cell behavior[35]. Specifically, substrate stiffness has been shown to play a role in cell adhesion[36,37] and morphology[37-39] and migration rate[37,40,41], growth[42], and differentiation[37,43-45]. Thus, we targeted global incorporation of the unnatural amino acid L-**Hag** periodically within an elastin-mimetic polypeptide derived from the consensus pentapeptide repeat (Val-Pro-Gly-Xaa-Gly).

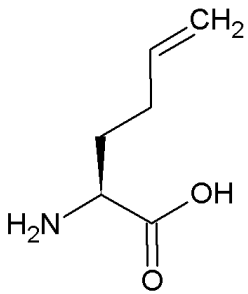
Herein, we describe the synthesis of a DNA concatemer encoding a highly repetitive elastin-mimetic polypeptide where methionine (Met) is encoded periodically within the elastin-mimetic polypeptide (Scheme 2) at position Xaa. A methionine-auxotrophic cell strain is utilized to direct overexpression of the elastin-mimetic polypeptide. Under selective pressure, where the growth medium lacks the natural amino acid, the unnatural amino acid L-**Hag** will serve as a substrate for the wild-type aminoacyl-tRNA synthetase permitting incorporation of the analogue in response to the Met codon. Encouraged by previous investigations in our lab[46], we anticipate that this residue-specific strategy will promote incorporation of the analogue within the elastin-mimetic polypeptide with a degree of biosynthetic substitution greater than what was previously reported[26-28,47,48]. Furthermore, the host strain will be cotransformed with a plasmid expressing an additional copy of the wild-type *E. coli* methionyl-tRNA synthetase (MetRS). By increasing the aaRS activity within the expression system, we may increase the overall protein yield of the analogue containing elastin-mimetic polymer. In addition, an amphiphilic (**AB**-type) diblock copolymer was constructed from the previously synthesized[49] elastic (**S1**) concatemer, which represents the hydrophilic (**A**) block, and the Hag-containing elastin-mimetic domain, which we hypothesize will exhibit a lower inverse temperature transition due to the nonpolar side chain of the analogue. Previous work in our lab[50] has demonstrated the ability of an amphiphilic diblock

copolymer derived from elastin-mimetic peptide sequences to undergo reversible, temperature-dependent segregation of the hydrophobic block in aqueous solution affording potentially biocompatible nanoparticles[50]. Incorporation of the alkene functionality within the hydrophobic block may permit both *physical* and *chemical* cross-linking. This amphiphilic diblock copolymer has the potential to capture features unique to both physical and chemical cross-linking schemes and subsequently to permit control over a wide range of chemical, mechanical, and biological properties afforded by protein-based multi-domain materials.

(a)



L-methionine (**Met**)



(*2S*)-2-amino-5-hexenoic acid (L-**Hag**)

(b)

```
                  Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
                  ↓
AG CTT GAA GAC GTT CCA GGT GTT GGC GTA CCG GGT GTA GGC GTA CCA GGC
     A CTT CTG CAA GGT CCA CAA CCG CAT GGC CCA CAT CCG CAT GGT CCG
HinD III Bbs I      ↑

Met Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
                                        ↓
ATG GGT GTA CCG GGT GTA GGC GTA CCG GGT GTT GGT GTT CCA AGA GAC GG
TAC CCA CAT GGC CCA CAT CCG CAT GGC CCA CAA CCA CAA GGT TCT CTG CCC TAG
                                                  ↑  Bsm B I   Bam H I
```

Scheme 1. (a) Chemical structures of the canonical amino acid (L-**Met**) and the methionine analogue (L-**Hag**) that were used as substrates for biosynthetic incorporation into the elastin-mimetic sequence, **elastin-M**, where the methionine codon represents the site of substitution with the analogue. (b) Oligonucleotide cassette encoding the **elastin-M** repeat sequence. The recognition sites for the relevant restrictions endonucleases, which are employed for generation of the DNA monomer, are highlighted. Arrows indicate the cleavage positions on the sense and anti-sense strand for the *Bbs* I and *Bsm*B I restriction endonucleases.

Experimental Methods

Materials

All chemical reagents were purchased from Fisher Scientific, Inc. (Pittsburgh, PA) or Sigma-Aldrich Corporation (St. Louis, MO), or VWR International, LLC (Radnor, PA) unless otherwise noted. Methionine analogue, (*2S*)-*2*-amino-5-hexenoic acid (L-**Hag)** was purchased from RSP Amino Acids LLC (Shirley, MA; RSP ID 0036). Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases were purchased from both New England Biolabs, Inc. (Beverly, MA) and Promega Corporation (Madison, WI), while antartic phosphatase, T4 DNA ligase, T4 polynucleotide kinase, DNA polymerase I large (Klenow) fragment, and deoxynucleotide solution mix (dNTPs) were purchased only from New England Biolabs, Inc. Shrimp alkaline phosphatase was obtained from Roche Applied Science (Indianapolis, IN). Platinum® *Pfx* DNA polymerase, plasmid pZErO®-1 and DNA DipStick™ were obtained from Invitrogen Corp. (Carlsbad, CA).

Plasmid pQE-80L was purchased from QIAGEN Inc.-USA (Valencia, CA) while plasmids pIL2 and pIL5 were synthesized by I-Lin Wu, a graduate student in the Conticello lab at Emory University. Plasmid pPROTetE.133 was obtained from BD Biosciences, Inc. (Palo Alto, CA) and pPROLarA.231 was obtained from Professor Rik Myers of the University of Miami. Plasmid pHEC1 was previously synthesized by Dr. Holly Carpenter in the Conticello lab at Emory University[51]. Plasmid DNA preparation and purification were performed using the Mini-Prep 24™ (MacConnell Research Corporation, San Diego, CA), the QIAfilter Plasmid Maxi Kit, QIAprep Spin Miniprep Kit, QIAquick PCR Purification Kit, and QIAwell 8 Ultra

Plasmid Kit (QIAGEN Inc., Valencia, CA), and the DNA Clean & Concentrator™ and Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA). Synthetic oligonucleotides were purchased from either Sigma-Genosys (a division of Sigma-Aldrich Corporation, St. Louis, MO) or Integrated DNA Technologies (Coralville, IA) and were used as received. The *E. coli* strain TOP10F´ was obtained from Invitrogen Corp. (Carlsbad, CA), XL10-Gold ultracompetent and BL21-Gold (DE3) from Stratagene (La Jolla, CA), and CAG18491[52,53] from the *E. coli* Genetic Stock Center at Yale University (New Haven, CT; CGSC#: 7464). New minimal medium (NMM) was prepared according to the protocol of Budisa et al.[23] with the exception that L-methionine was not added to the medium prior to cell culture.

General Methods

The molecular biology techniques utilized here, including expression cloning, polymerase chain reaction, gel electrophoresis, and growth and induction of bacterial cultures were adapted from a standard molecular cloning manual[54] or the protocol supplied by manufacturer, unless otherwise noted. Reagents for the manipulation of DNA, bacteria, and recombinant proteins were sterilized by either autoclave or passage through a syringe filter (0.2 μm cellulose membrane) or vacuum filter unit (standard polyethersulfone (PES) membrane) available from VWR International, LLC (Radnor, PA). Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. The concentration of DNA solutions was measured by an Ultrospec 3000 UV/Visible Spectrophotometer (Pharmacia Biotech, Cambridge, UK). Site-directed mutagenesis was performed using Stratagene's (La Jolla, CA) Quick-Change mutagenesis technique from gene-specific oligonucleotide primers. Polymerase chain reaction (PCR) was carried out using a GeneAmp 2400 Thermal Cycler (PerkinElmer Inc. Waltham, MA)

and MJ Mini™ Gradient Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules, CA). Over the course of these studies, automated DNA sequencing was performed at each of the following facilities: Beckman Coulter Genomics (Danvers, MA) on a Perkin-Elmer ABI Prism 377 DNA sequencer and GENEWIZ, Inc. (South Plainfield, NJ) on an Applied Biosystems 3730xl DNA analyzer. Agarose gel electrophoresis images were captured using a Kodak DC-120 digital camera or Kodak Gel Logic 112 imaging system from Carestream Health, Inc. (Rochester, NY). Protein electrophoresis was performed using 10-15% gradient discontinuous SDS PAGE pre-cast gels with a PhastSystem and visualized by silver staining using PlusOne Silver Stain Kit for proteins from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ). Amino acid compositional analyses were performed at the W. M. Keck Foundation Biotechnology Resource Laboratory of Yale University (New Haven, CT) using the Hitachi L-8900 PH amino acid analyzer.

Construction of the elastin-M gene

The **elastin-M** multimer, which encodes a highly repetitive, elastin-mimetic polymer, was constructed using a previously reported concatemerization strategy[49]. The procedure for DNA cassette concatemerization was carried out as described in detail in Chapter 2 of this dissertation, unless otherwise noted. The DNA oligonucleotide primers encoding the sense and anti-sense monomer sequence (Scheme 1) were chemically synthesized and annealed to produce duplex monomer DNA. Synthesis of the DNA duplex was visualized by agarose gel electrophoresis. The double-stranded DNA fragment was purified via ethanol precipitation with Pellet Paint® co-precipitant and addition of 5'-phosphates to the double stranded monomer DNA was performed using T4 polynucleotide kinase. The reaction was extracted with phenol/chloroform/isoamyl alcohol (25:24:1) to remove the kinase enzyme. Ethanol precipitation

was used to isolate the phosphorylated DNA duplexes and reduce the salt concentration. Approximately 400-500 ng of phosphorylated, DNA monomer was isolated and stored in 50 μL buffer EB at –20 ºC.

A ligation between the monomer DNA (~320 ng) and *Bam*H I/*Hin*d III-digested pZErO®-1 plasmid (~150 ng) was performed in a 3:1 molar ratio in the presence of 1 μL T4 DNA ligase enzyme (400 units) with incubation at 16 ºC for 30 min. An aliquot (2 μL) of the ligation mixture was used to transform electrocompetent cells of *E. coli* strain TOP10F´ (40 μL). The cells were recovered in 1 mL SOC rich media for 1 h at 37 ºC with shaking at 225 rpm. Aliquots (150 μL) of each transformation suspension were spread onto low salt LB agar media supplemented with 1 mM IPTG and Zeocin™ (50 μg/mL) for antibiotic selection. The plates were incubated at 37 °C for 12 to 14 h. Twelve single colonies were picked from the plate and used to inoculate separate culture tubes of low salt LB media (5 mL) supplemented with Zeocin™, which were grown overnight at 37 ºC in a rotating drum. The plasmid DNA was isolated using the QIAprep Spin Miniprep Kit for each of the samples. Recombinant plasmids were screened by double digestion with *Bam*H I/*Hin*d III to identify the elastin-M monomer gene. The DNA digestion products were visualized via agarose gel electrophoresis (4% NuSieve® GTG® agarose, 1X TBE buffer). The sequence of the elastin-M monomer gene was confirmed by automated DNA sequencing analysis using oligonucleotide primers M13 reverse and M13 (-20) forward primers (Appendix 1), which bind upstream and downstream of the target sequence respectively. All of the plasmid DNA samples contained the correct gene of interest. Recombinant plasmids containing the correct insert sequence of the elastin-M DNA monomer flanked by *Bbs* I and *Bsm*B I restriction endonuclease recognition sites were designated **pMAP18** and stored at –20 °C.

A large amount of the monomer was needed for DNA cassette concatemerization. The **pMAP18** plasmid DNA was isolated from four-2.8 L flasks containing 500 mL of low salt LB media supplemented with Zeocin™ (50 μg/mL) using the QIAfilter Plasmid Maxi Kit (approximately 1.5 mg yield). The elastin-M DNA monomer was liberated from the cloning by sequential digestion with *Bbs* I and *Bsm*B I. The digested **pMAP18** plasmid DNA was loaded onto a preparative 2% NuSieve® GTG® agarose gel (1X TBE buffer) in order to separate the pZErO®-1 vector from the elastin-M monomer insert. The band corresponding to the 75 bp elastin-M monomer 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.

A ligation between monomers (multimerization) was prepared by incubating the isolated monomer DNA (3 μg; 40 μL) in the presence of T4 DNA ligase at 16 °C for 16 h. The library of DNA concatemers was fractionated via agarose gel electrophoresis. The region corresponding to the desired multimer size range (~2500 to 1000 bp) was excised from the gel. The DNA was isolated using the Zymoclean™ Gel DNA Recovery Kit and quantified using DNA DipStick™. A *Bbs* I-digested **pMAP18** acceptor plasmid (~1.2 μg) was dephosphorylated at the 5'-ends with antartic phosphatase (10 units/μg) in 1X antartic phosphatase buffer at 37 °C for 30 min followed by heat inactivation at 65 °C for 5 min. The pool of DNA multimers of size 1000 to 2500 bp was ligated back into the *Bbs* I -digested **pMAP18** acceptor plasmid in a 1:1 ratio (w/w) via incubation with 1 μL T4 DNA ligase (400 units) and 1X T4 DNA ligase buffer at 16 °C for 16 h. An aliquot of the ligation (2 μL) was transformed into electrocompetent TOP10F´ cells. The recovery mixture (150 μL) was spread onto low salt LB agar media supplemented with 1 mM

IPTG and Zeocin™ (50 μg/mL). Twenty-four single colonies were selected and DNA, representing the elastin-M DNA concatemers in pZErO®-1, was isolated using the automated Mini-Prep 24™. The recombinant plasmids were screened via double digestion with *Bam*H I/*Hin*d III and analyzed via agarose gel electrophoresis (1% agarose, 0.5X TBE buffer) for the presence of multimer inserts.

From the pool of transformants, a colony was isolated containing a recombinant plasmid with a concatemer insert in the desired size range (~1500 bp). The concatemer clone was used to transform fresh electrocompetent TOP10F' cells and plasmid DNA was isolated from the bacteria using the QIAprep Spin Miniprep Kit. The sequence of the elastin-M multimer gene was confirmed by automated DNA sequencing analysis using oligonucleotide primers M13 reverse and M13 (-20) forward (Appendix 1), which bind upstream and downstream of the target sequence respectively. The isolated plasmid, which contains the elastin-M concatemer gene, was designated **pMAP20** (Figure 3) and stored at –20 °C.

Assembly of the diblock gene fusion

A second plasmid was generated that encodes a diblock (**AB**-block) copolymer where the previously synthesized hydrophilic, elastin (**S1**) block (*see Chapter 2*) is encoded at the N-terminus and the elastin-M concatemer at the C-terminus. Since the antibiotic resistance gene for the **S1** and elastin-M corresponding cloning plasmids pZErO®-2.1 and pZErO®-1 differ, use of the directional cloning strategy required that both DNA concatemer cassettes be encoded in the same cloning plasmid. Therefore, the **S1** multimer was removed from the pZErO®-2.1 vector via double digestion of **pSCP3.11** (5 μg) with *Bam*H I and *Hin*d III (5 units/μg of each enzyme) in 1X MULTI-CORE™ Buffer (Promega) at 37 °C for 4 h. Approximately 1 μg of multimer insert

was isolated via agarose gel electrophoresis (1% agarose, 0.5X TBE buffer) and purified using the Zymoclean™ Gel DNA Recovery Kit. A ligation between the **S1** multimer (100 ng) and *Bam*H I/*Hin*d III-digested pZErO®-1 plasmid was performed in a 3:1 molar ratio in the presence of 1 μL T4 DNA ligase enzyme (400 units) and 1X T4 DNA ligase buffer at 16 °C for 30 min. An aliquot (2 μL) of the ligation mixture was used to transform electrocompetent TOP10F' cells. The recovery mixture (150 μL) was spread on low salt LB agar media supplemented with 1 mM IPTG and Zeocin™ (50 μg/mL). Twelve single colonies were picked from the plate and used to inoculate separate culture tubes of low salt LB media (5 mL) supplemented with Zeocin™, which were incubated overnight at 37 ºC in a rotating drum. The plasmid DNA was isolated using the QIAprep Spin Miniprep Kit for each of the samples. Recombinant plasmids were screened by double digestion with *Bam*H I/*Hin*d III to identify the **S1** multimer gene in the plasmid. Two colonies representing the most promising inserts were cultured and DNA was isolated from the bacteria using the QIAprep Spin Miniprep Kit. The purified recombinant plasmid clones, designated **pMAP19** (Figure 3), were transformed into electrocompetent TOP10F' cells, DNA was isolated using the QIAprep Spin Miniprep Kit, and stored at –20 °C.

Directional assembly of the concatemers involved the use of the type IIs restriction endonucleases and a third enzyme *Xma* I, which cleaves within the antibiotic resistance gene of the pZErO®-1 plasmid. A ligation was performed between *Bsm*B I/*Xma* I-digested **pMAP19** (~80 ng) and *Bbs* I/*Xma* I-digested **pMAP20** (~100 ng) in a 1:1 molar ratio with 1 μL T4 DNA ligase (400 units) and 1X T4 DNA ligase buffer in a reaction volume of 10 μL at 16 °C for 16 h. An aliquot (2 μL) of the ligation mixture was used to transform electrocompetent TOP10F' cells, which were plated onto low salt LB agar media supplemented with 1 mM IPTG and Zeocin™

(50 μg/mL). The plates were incubated at 37 °C for 12 to 14 h. Four single colonies were picked from each plate and used to inoculate culture tubes of low salt LB media (5 mL) supplemented with Zeocin™ grown overnight at 37 °C in a rotating drum. The recombinant plasmids were screened by digestion with *Bam*H I, *Bbs* I, and *Dra* I and DNA fragments were analyzed by agarose gel electrophoresis (1% agarose, 0.5X TBE buffer) for the presence of the diblock insert. The sequence of the purified plasmid DNA, designated **pMAP23** (Figure 5), containing the elastin-M (AB) diblock gene fusion, was confirmed by automated DNA sequencing analysis with the M13 primers. Correctly sequenced clones were identified by the presence of a gene fusion of the **S1** and elastin-M multimers flanked by *Bbs* I and *Bsm*B I restriction endonuclease recognition sites, respectively. The purified plasmid DNA was used to transform electrocompetent TOP10F' cells and XL10-Gold Ultracompetent cells, the latter of which are suitable for cloning of large plasmids. Aliquots (150 μL) of each TOP10F' and XL10-Gold Ultracompetent recovery mixture were spread onto low salt LB agar media supplemented with 1 mM IPTG and Zeocin™ (50 μg/mL) or supplemented with Zeocin™, chloramphenicol (35 μg/mL) and tetracycline (10 μg/mL), respectively. The plates were incubated at 37 °C for 12 to 14 h. Single colonies of from each transformation plate were used to inoculate LB media (5 mL) supplemented with Zeocin™ alone or Zeocin™ and chloramphenicol. The cells were cultured at 37 °C in a rotating drum to an OD_{600nm} of 0.8 to 1.0. Glycerol stocks of the cell strains were prepared from 800 μL of cell culture and 200 μL of sterile 80% glycerol and stored at –80 °C. Next, an acceptor plasmid suitable for protein expression was required to conduct high-level expression of **elastin-M** and **elastin-M** (**AB**) variants in a bacterial host.

Expression plasmid construction

The genes encoding the elastin-mimetic polypeptides were cloned into the commercially available pQE-80L expression vector. Use of the seamless cloning strategy required insertion of an adaptor sequence containing recognition sites for type IIs restriction endonucleases. The modified expression plasmids **pIL2** and **pIL5** were synthesized by I-Lin Wu, a graduate student in the Conticello lab, and their synthesis is summarized as follows. An adaptor sequence (Scheme 2) was designed for cloning of the elastin-M concatemer cassettes in the pQE-80L expression plasmid. The single-stranded oligonucleotides corresponding to the sense and anti-sense strands of the **pQE Adaptor** were chemically synthesized. Annealing was carried out by dissolving the primers in sterile ddH_2O to a final concentration of 0.5 μg/μL. Twenty microliter aliquots of each of the two primers were mixed together with 4 μL 5M NaCl, 4 μL 1M $MgCl_2$, and 152 μL sterile ddH_2O. By gradually decreasing the temperature of the reactions from 99 °C to 30 °C (decreased by 1 °C every 5 min), the DNA strands were annealed together. The 3' recessed ends of the annealed, double stranded DNA were filled in with the DNA polymerase I, large (Klenow) fragment (4 μL; 20 units) in 1X NEBuffer 2 and 33 μM dNTP solution at 25 °C for 2 h. The DNA duplex was purified using the QIAquick PCR purification Kit. The adaptor and commercially available pQE-80L were digested with *Eco* RI and *Hin*d III (40 units of each enzyme) and separated via 4% NuSieve®GTG® agarose (1X TBE buffer) and 1% agarose (0.5X TBE buffer) electrophoresis, respectively. The adaptor and vector were purified using the Zymoclean™ Gel DNA Recovery Kit. A ligation reaction between the digested adaptor and pQE-80L plasmid was performed with 0.5 μL T4 DNA ligase (200 units) and 1X T4 DNA ligase buffer in a 10 μL reaction volume at 16 °C for 12-14 h. An aliquot (2 μL) of the ligation reaction

was then used to transform electrocompetent TOP10F' cells (40 μL) and the recovery mixture (100 μL) was spread on LB agar media supplemented with ampicillin (100 μg/mL). Six single colonies were used to inoculate LB media (5 mL) supplemented with ampicillin and grown overnight at 37 °C in a rotating drum. The plasmid DNA, designated **pIL2**, was purified with QIAprep Spin Miniprep Kit. The sequence of the recombinant plasmid clones was confirmed by automated DNA sequencing using primers designed to bind the pQE-80L plasmid upstream and downstream of the inserted adaptor sequence (Appendix 1).

For subsequent cloning of the elastin concatemers, it was necessary to remove a *Bsa* I restriction endonuclease recognition site from the modified pQE-80L expression plasmid, **pIL2**. 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 and produced plasmid **pIL5**. The forward mutagenesis primer, pQE80LBsaI-F (5'-CAGTGCTGCAATGATA CCGCGAGTGCCACGCTCACCGGCTCCAGATT-3'), and reverse mutagenesis primer, pQE80LBsaI-R (5'-AATCTGGAGCCGGTGAGCGTGGCACTCGCGGTATCATTGCAGCAC TG-3'), were suspended in ddH_2O at a concentration of 100 μM. The **pIL2** template plasmid was amplified using the Platinum® *Pfx* DNA polymerase (1 μL) in thin-walled PCR tubes containing a reaction volume of 50 μL (1X *Pfx* Amplification Buffer, 0.3 μM of each mutagenic primer, 0.3 mM each dNTP, 1 mM $MgSO_4$). The tubes were placed in the heat block of a thermal cycler and subjected to the following amplification regime: 95 °C for 2 min; 30 cycles of 95 °C for 1 min, 55 °C for 1 min, 68 °C for 2 min; 68 °C for 7 min; and final hold (4 °C) for short-term storage. PCR products were analyzed by 1% agarose gel electrophoresis (0.5X TBE buffer) and automated DNA sequencing using primers designed to bind upstream and downstream of the

pQE-80L multiple cloning site.

Cloning of elastin-M genes into expression plasmid

The pZErO®-1 plasmids, **pMAP20** and **pMAP23**, containing the corresponding elastin-M concatemer and elastin-M (AB) diblock gene fusion (~10 μg of each) were sequentially digested with *Bbs* I and *Bsm*B I to isolate the multimer genes. The **pIL5** plasmid was digested with *Bsa* I to generate the complimentary sticky ends necessary for seamless cloning of the elastin-M genes into the adaptor sequence of the modified pQE-80L expression plasmid. The digested **pIL5** DNA was also dephosphorylated with antartic phosphatase to reduce the incidence of false positive ligations. Ligation reactions were incubated at 16 °C for 12 h to covalently link the *Bsa* I-digested **pIL5** to the elastin-M inserts. An aliquot (2 μL) of each of the ligations was transformed into electrocompetent cells of TOP10F'. The recovery mixture (100 μL) was spread on LB agar media supplemented with ampicillin (100 μg/mL). Eight single colonies of were picked from each plate for screening. The recombinant plasmid DNA was isolated from the cultured colonies under ampicillin selection using the QIAprep Spin Miniprep Kit. The DNA was screened via double digestion with *Eco*R I and *Hind* III for the presence of the elastin-M monoblock and diblock (AB-type) concatemer inserts. Automated DNA sequencing analysis using primers designed to bind the pQE-80L plasmid upstream and downstream of the adaptor sequence confirmed the sequence of the expression plasmids. Plasmid **pMAP36** includes the pQE-80L plasmid with the 1500 bp elastin-M gene encoding the **elastin-M** polymer, while **pMAP34** includes the pQE-80L plasmid with the 3000 bp diblock gene fusion encoding the **elastin-M** (**AB**) copolymer. Glycerol stocks of the **pMAP36** and **pMAP34** TOP10F' transformants were prepared from 800 μL of overnight LB cell culture (5 mL) and 200

μL 80% glycerol and stored at –80 °C.

Synthetase plasmid construction

Synthesis of the acceptor plasmid **pME1**, for expression of the aminoacyl-tRNA synthetase genes, was previously reported[46]. Briefly, **pME1** 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, generating the hybrid pPROTet/Lar plasmid. The plasmid was utilized for subsequent cloning of the gene encoding the wild-type methionyl-tRNA synthetase (MetRS). A recombinant copy of the *Escherichia coli* MetRS gene, *metG* (Appendix 2), was amplified from the genomic DNA using gene specific primers, metG-F (5'-AGGTTGGTACCATGACTCAAGTCGCGAAGAAAATTCTGGTGA-3') and metG-R (GGAGCTCTAGATTATTTCACCTGATGACCCGGTTTAGCACCGGCATCCG GG-3'), designed to introduce *Kpn* I and *Xba* I restriction endonuclease sites at the 5' and 3' termini of the amplified gene, respectively. A culture tube of LB media (2 mL) was inoculated with *E. coli* strain TOP10F' and genomic DNA was isolated by the boiling lysis miniprep method[55]. The PCR reaction (1x *Pfx* Amplification buffer, 0.3 mM dNTP, 1 mM MgSO4, 0.3 μM each primer, and 2.5 units Platinum® *Pfx* DNA Polymerase) was subjected to the following amplification regime: 95 °C for 2 min; 30 cycles of 94 °C for 15 sec, 55 °C for 30 sec, 68 °C for 2.5 min; 68 °C for 7 min; and final hold at 4 °C for short-term storage. The PCR amplified *metG* gene was purified from the reaction mixture using the DNA Clean and Concentrator™, yielding approximately 2 μg of DNA. After amplification, the gene product (2 μg) was double digested with *Kpn* I (25 units) and *Xba* I (30 units) in 1X NEBuffer 2 and 1X BSA at 37 °C for 4 h. The

DNA was purified via agarose gel electrophoresis (1% agarose, 0.5X TBE buffer) using the Zymoclean™ Gel DNA Recovery Kit.

The acceptor expression plasmid **pME1** was double digested with *Kpn* I and *Xba* I, dephosphorylated with shrimp alkaline phosphatase, and purified using the Zymoclean™ Gel DNA Recovery Kit. A ligation reaction was prepared between the digested, dephosphorylated pPROTet/Lar expression plasmid (~100 ng) and the digested PCR product (~300 ng) with 1 μL T4 DNA ligase and 1X T4 DNA ligase buffer at 16 °C for 16 h. An aliquot (4 μL) of the ligation reaction was used to transform electrocompetent TOP10F' cells and the recovery mixture (150 μL) was spread on LB agar media supplemented with chloramphenicol (35 μg/mL). Six single colonies were selected for culturing in LB media (5 mL) supplemented with chloramphenicol and the recombinant plasmid DNA was purified using the QIAprep Spin Miniprep Kit. The presence of an insert of size concomitant with insertion of the MetRS cassette was visualized by 1% agarose gel electrophoresis (0.5X TBE buffer). Automated DNA sequencing analysis with primers designed to bind the pPROTet/Lar plasmid upstream and downstream, respectively, of the multiple cloning site confirmed the sequence. The new plasmid **pMetRS** contains a copy of the *E. coli* wild-type metG gene, encoding the methionyl-tRNA synthetase (MetRS), under control of the $P_L Tet$ promoter. Glycerol stocks of TOP10F' transformants containing correctly sequenced clones of **pMetRS** were prepared from 800 μL of overnight LB cell culture (5 mL) and 200 μL 80% glycerol and stored at –80 °C.

Bacterial growth and protein expression

Small scale expression in CAG18491 strain

The plasmids **pMAP36** and **pMAP34**, which encode the **elastin-M** and **elastin-M** (**AB**)

polymers, respectively, were cotransformed with the **pMetRS** plasmid (encoding the wild-type *E. coli* methionyl-tRNA synthetase) or with the plasmid **pHEC1** (a modified pPROTet/Lar plasmid lacking the MetRS cassette) into the methionine auxotroph CAG18491 [F-, λ-, *rph-1*, *metEo-3079::Tn10*]. The cells were plated on LB agar media supplemented with ampicillin (100 μg/mL), chloramphenicol (35 μg/mL), and tetracycline (10 μg/mL). Glycerol stocks of the expression strains employed for these studies were prepared and stored at –80 °C.

Five milliliters of LB media, supplemented with ampicillin, chloramphenicol, and tetracycline as required for plasmid maintenance, were inoculated with single colonies of the CAG18491 expression strains and grown overnight at 37 °C on a rotating drum. One milliliter (2% of the final culture volume) of the overnight cultures was used to inoculate new minimal media (NMM; 7.5 mM $(NH_4)_2SO_4$, 8.5 mM NaCl, 55 mM KH_2PO_4, 100 mM K_2HPO_4, 1 mM $MgSO_4$, 20 mM glucose, 1 mM $CaCl_2$, 10 mg/mL thiamine hydrochloride, 10 mg/mL biotin in PBS, and 50 μg/mL of each of the 20 canonical amino acids except methionine) supplemented with antibiotics and 0.4 mM L-methionine (prepared as a 200 mM stock solution in 1X phosphate buffered saline, PBS at pH ~7.0) for a final culture volume of 50 mL. The fifty-milliliter cultures were incubated at 37 ºC with shaking at 225 rpm to an OD_{600nm} of 0.8 to 1.0 over a period of approximately 2 to 3 h. Once the cells reached log phase growth, the intact cells were harvested by centrifugation at 4 °C, 4000 x *g* for 10 min. The supernatant was discarded, and the cell pellets were gently suspended in 10 mL aqueous, sterile ice-cold 0.9 % NaCl solution. The cell suspensions were centrifuged at 4 °C, 4000 x *g* for 10 min. The cell pellets were washed once more in the isotonic sodium chloride solution. The supernatant was discarded, and the cell pellets were finally suspended in 40 mL of NMM without addition of antibiotics or

methionine. The cell suspension was then divided into four-10 mL aliquots and incubated at 37 °C for 30 min with shaking at 225 rpm to deplete the remaining cellular levels of methionine.

The methionine analogue, (2*S*)-2-amino-5-hexenoic acid (L-**Hag)**, was prepared as a 200 mM stock solution in 1X PBS with titration of 1 M NaOH until soluble and stored at 4 °C. Aliquots of the methionine and the L-**Hag** stock solutions were added to the 10 mL cell cultures for final concentrations ranging from 0.5 mM to 1 mM. An aliquot (10 μL) of a 1 M IPTG solution (prepared by dissolving 596 mg IPTG in 2.5 mL ddH_2O followed by filter sterilization) was added to the cultures to a final concentration of 1 mM to induce elastin-mimetic protein expression. The cultures were incubated at 37 °C with shaking at 175 rpm for 3 h after induction. One-milliliter aliquots of the cell cultures were collected after 3 h. The aliquots were centrifuged at 4 °C for 5 min at 6,300 rpm, the supernatant was discarded, and the samples were suspended in 30 μL sterile ddH_2O and stored at –20 °C. The cell culture samples were analyzed by SDS PAGE for evidence of protein synthesis. Briefly, whole cell lysates were prepared by mixing aliquots (5 μL) of the cell culture samples with 12.5 μL 2X SDS gel-loading buffer (100 mM Tris-Cl, pH 6.8; 4% (w/v) SDS, electrophoresis grade; 0.2% (v/v) bromophenol blue; 20% (v/v) glycerol), 2.5 μL 1 M DTT, and sterile ddH_2O to a final volume of 25 μL and boiling mixture at 100 °C for 5 min. Samples were cooled on ice for 10 min prior to electrophoresis.

Large-scale expression

Plasmids **pMAP36** and **pMAP34** (pQE-80L plasmids encoding the respective **elastin-M** and **elastin-M** (**AB**) polymers) were cotransformed with the **pMetRS** plasmid (containing a recombinant, wild-type *E. coli* methionyl-tRNA synthetase gene) into the CAG18491 strain and plated on LB agar media supplemented with ampicillin (100 μg/mL), chloramphenicol (35

μg/mL), and tetracycline (10 μg/mL). Twenty milliliters of LB media supplemented with ampicillin (100 μg/mL), chloramphenicol (35 μg/mL), and tetracycline (10 μg/mL) were inoculated with single colonies of CAG18491 cells, harboring the appropriate plasmids, and grown overnight at 37 °C with shaking at 225 rpm. The overnight seed cultures were washed with 20 mL fresh LB media and suspended in 20 mL fresh LB media.

The cell suspensions were divided into 2-10 mL aliquots and each was transferred to a 2.8 L flask containing 490 mL NMM (supplemented with ampicillin (100 μg/mL), chloramphenicol (35 μg/mL), and tetracycline (10 μg/mL) for a total culture volume of 500 mL. In this way, two flasks were used for elastin-mimetic biosynthesis from a 1000 mL (1 L) total culture volume for each cell strain. The large cultures were allowed to grow at 37 ºC to an OD_{600nm} of 0.8 to 1.0 over a period of approximately 2 to 3 h. Once the cells reached log phase growth, the intact cells were harvested by centrifugation of the 500 mL cultures at 4 °C, 4000 x *g* for 10 min. The supernatant was discarded, and the cell pellets were washed with an aqueous solution of ice-cold, sterile 0.9 % NaCl (2x 100 mL). The cell pellets were finally suspended in 500 mL of sterile NMM lacking methionine, but supplemented with antibiotics (ampicillin (50 μg/mL), chloramphenicol (35 μg/mL), and tetracycline (2.5 μg/mL)) at reduced concentrations to maximize protein expression. The cell cultures were incubated at 37 °C for 30 min with shaking at 225 rpm to deplete the remaining cellular levels of methionine.

The methionine analogue, (2*S*)-2-amino-5-hexenoic acid (L-**Hag)**, was prepared as a 200 mM stock solution in PBS with titration of 1 M NaOH for a final pH of ~12 and stored at 4 °C. An aliquot of 1.25 mL of methionine or the L-**Hag** stock solutions was added to the 500 mL cell cultures for a final concentration of 0.5 mM. An aliquot (500 μL) of a 1 M IPTG solution was

added to the cultures to a final concentration of 1 mM to induce elastin-mimetic protein expression. The cultures were incubated at 37 °C with shaking at 175 rpm for 3 h after induction. The remaining cells were harvested by centrifugation at 4 °C, 4,000 x g for 20 min. Cell pellets of each 500 mL culture from the same expression strain were combined and suspended in 50 mL lysis buffer (50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and stored at –80 °C.

Purification of elastin-M polymers

The frozen cells were lysed by three freeze/thaw cycles at –80 °C and ambient temperature, respectively. Lysozyme was added to a final concentration of 2 mg/mL (stock solution of 10 mg/mL) and 1X Protease Inhibitor Cocktail Set I (Calbiochem®, a brand of EMD Chemicals; Gibbstown, NJ)) was added to the lysate. The mixture was incubated at room temperature for 30 min with shaking. Benzonase® nuclease (25 units/mL final concentration) and 1 mM $MgCl_2$ (stock solution of 1 M) were then added, and the cell lysis mixture was incubated with shaking at ambient temperature for 30 min. The solutions were incubated at 4 °C for 36 h.

The cell lysate was centrifuged at 20,000 x *g* for 40 min at 4 °C to pellet the cellular debris. 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 while some residual protein remained in the pellet. The elastin-mimetic polymers were then purified via inverse transition cycling according to a protocol adapted from Meyer and Chilkoti[56]. Five molar NaCl (~25 mL) was added to the 50 mL of soluble extracts for a final molarity of 2 M, which includes the 300 mM NaCl present in the lysis buffer. The **elastin-M** and **elastin-M** (**AB**) polymer solutions were incubated at 33 °C and 37 °C, respectively, for 45 to 60 min until precipitation of the elastin mimetic-polymers was

observed as evidenced by increased turbidity of the solution. The solutions were centrifuged at 33 °C and 37 °C, 10,000 x *g* for 15 min to pellet the precipitated proteins ('hot-spin'). The protein pellet was placed on ice and suspended in ice-cold (0 °C), sterile lysis buffer (~42 mL) and 1X Protease Inhibitor Cocktail Set I for 30 to 45 min. A second cycle of the centrifugation cycle was repeated beginning with centrifugation at 4 °C, 30,000 x *g* for 15 min ('cold-spin'). Following precipitation at 33/37 °C, the resulting protein pellets were suspended in ice-cold (0 °C), sterile lysis buffer (~35 mL) and 1X Protease Inhibitor Cocktail Set I for 30 to 45 min. The suspended protein pellet was then stored overnight at 4 °C with gentle shaking. Additional centrifugation cycles were repeated until no insoluble impurities were observed following centrifugation of the protein solution at 4 °C. Following the final hot-spin at 33 °C and 37 °C, the resulting protein pellets were suspended in ice-cold (0 °C), sterile lysis buffer (~10-15 mL), without addition of protease inhibitor cocktail, and incubated at 4 °C with shaking overnight.

The **elastin-M** and **elastin-M** (**AB**) polymer solutions were dialyzed at 4 °C using the SnakeSkin® Pleated Dialysis Tubing with a molecular weight cut-off of 10 kDa (Pierce Protein Research Products, Thermo Fisher Scientific Inc., Rockford, IL) against distilled, deionized water (5 x 4 L) to remove excess salts. The protein solutions were further purified via sterilization at 4 °C through a 5.0 μM Durapore® membrane filter (Millipore; Billerica, MA) using a Fisherbrand® glass microanalysis filter holder assembly. Lyophilization of the filtered dialysate produced the elastin-mimetic polypeptides as white spongy solids. The lyophilized proteins were analyzed by 10-15% gradient discontinuous SDS PAGE (PhastSystem) and protein bands were visualized by silver staining (PlusOne Silver Stain kit).

Physical and analytical measurements

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

MALDI-TOF MS experiments were performed on an Applied Biosystems® Voyager™ System 428 mass spectrometer (Life Technologies Corporation; Carlsbad, CA) in the positive ion linear mode. The matrices, sinapinic acid (3-(4-hydroxy-3, 5-dimethoxyphenyl)prop-2-enoic acid) and ferulic acid were used at a concentration of 20 mg/mL in a 75% acetonitrile: 0.2 % trifluoroacetic acid or 75% acetonitrile: 0.2% formic acid solution, respectively. The protein solution (~100 pmol/μL) was prepared by dissolving 0.5 mg to 1 mg of polymer in 250 μL sterile ddH_2O at 4 °C. One-microliter of the protein solution was mixed with the matrix solution in a ratio of 1:10 and vortexed gently for a few seconds. Two microliter aliquots of the mixture were spotted on a stainless steel, 100-position flat sample plate and dried in air. The Sequazyme™ bovine serum albumin (BSA) standard (Applied Biosystems Inc.; Carlsbad, CA) was used as a standard for external calibration. An aliquot (1 μL) of the suspended standard solution was prepared in 30 μL sinapinic acid matrix solution and spotted on the plate. To minimize precipitation of the proteins, special precautions were taken during spotting of the plate. Firstly, the elastin-mimetic protein samples were kept on ice immediately prior to mixing with the matrix and spotting of the grid. Second, after spotting the grid, the plate was covered in an empty Petri plate and placed at 4 °C while spots dried. All spectra were taken in linear positive ion mode.

Differential scanning calorimetry

Differential scanning calorimetry (DSC) was performed using the ultra-sensitive VP-

DSC instrument from MicroCal, LLC (Northampton, MA). Dilute solutions of the **elastin-M** monoblock polymer samples (0.75 to 1.0 mg/mL) were prepared at 4 °C in sterile ddH_2O and 40 mM ethanoic (acetic) acid or 40 mM NaOH. Samples were degassed under dynamic vacuum 4 °C for 30 min. The thermal transition data was recorded over a temperature range from 4 °C to 60 °C at a scan rate of 1 °C/min with a pre-scan thermostat of 15 min. Reversibility was tested by cooling and re-scanning of samples *in situ*. Data were processed and analyzed using the program ORIGIN (MicroCal, LLC, Northampton, MA).

Nuclear magnetic resonance (NMR) spectroscopy

Solution NMR spectra of the **elastin-M** monoblock polymers were acquired using a Varian UNITY Plus 600 instrument (600 Mhz ^{1}H) from Agilent Technologies (Santa Clara, CA). The ^{1}H-NMR spectra were collected at 4 °C on samples consisting of 10 to 15 mg of protein dissolved in 0.7 mL deuterium oxide (D, 99%) from Cambridge Isotope Laboratories, Inc. (Andover, MA). Chemical shifts for the ^{1}H nuclei were referenced and reported relative to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS)[57]. The NMR data were processed using the ACD/NMR Processor Academic Edition from ACD/Labs (Ontario, Canada).

Temperature-dependent turbidity

The thermal transitions of the **elastin-M** (**AB**) were identified from temperature turbidity profiles. Measurements were performed using a JASCO V-530 UV/Visible spectrophotometer equipped with a programmable water-controlled Peltier cell and a JASCO ETC-505T temperature controller (Jasco, Inc.; Easton, MD). Dilute solutions of the elastin-mimetic diblock copolymer samples (0.5 to 1.0 mg/mL) were prepared at 4 °C and analyzed in a quartz cuvette with a 10 mm path length (Hellma Analytics; Müllheim, Germany). The absorbance of the

copolymer solutions was monitored over a temperature range of 4 °C to 80 °C at a wavelength of 256 nm with a ramp rate of 1 °C/min. Rescan of the samples was performed *in situ* following a 15 min pre-scan thermostat. Spectra were recorded and plotted using the Temperature-Scan (Melting) analysis feature of the JASCO Spectra Manager software package.

Transmission electron microscopy

Dilute aqueous solutions (1 mg/mL) of **elastin-Hag** (**AB**) diblock copolymer were prepared at neutral pH and low pH (40 mM acetic acid). The solutions were incubated at 37 °C for approximately 1 h to induce hydrophobic assembly of the polypeptide chains and to generate a coacervate. Aliquots (4 μL) of the polymer solutions were then deposited onto the surface of a 200 mesh formvar-carbon film-coated copper grids at room temperature and allowed to set for 60 s. Excess solvent was wicked away and 4 μL of a 1:1 mixture of the negative stains NanoVan® (methylamine vanadate) and Nano-W® (methylamine tungstate) from Nanoprobes (Yaphank, NY) was applied to the grid. After incubating at room temperature for 60 s, excess stain was wicked away and the grids were dried under vacuum. The samples were examined with a Hitachi H-7500 TEM at a 70 kV accelerating voltage.

Results and Discussion

The elastin-mimetic polypeptide **elastin-M** (Scheme 1) was designed as an artificial polymer for biosynthetic replacement of methionine by the analogue L-**Hag** using an IPTG-inducible expression system. As a result, the terminal vinyl groups are incorporated periodically along the polypeptide chain. The repetitive domain of **elastin-M** is derived from the pentapeptide repeat of elastin (Val-Pro-Gly-Xaa-Gly). Annealing of the **elastin-M** synthetic oligonucleotides assembled the duplex DNA monomer. The presence of the *Hin*d III and *Bam*H I restriction sites at the 5' and 3' termini respectively, facilitates insertion of the DNA monomer into the multiple cloning region of the plasmid pZErO®-1 (Figure 1). The pZErO®-1 cloning vector was selected for more recent DNA concatemerization protocols because the presence of the smaller Zeocin™ resistance gene reduces the overall vector size for increased transformation and ligation efficiencies. The seamless cloning strategy utilizing type IIs restriction endonucleases was employed to liberate the DNA monomer from the cloning plasmid for subsequent concatemerization. The library of concatemers was fractionated via agarose gel electrophoresis generating a ladder of bands where the size of individual concatemers differs by the unit of a single DNA monomer, approximately 75 base pairs. Concatemers in the desired size range of 1000 to 2500 bp were recloned into the original **elastin-M** monomer plasmid and the recombinant clones were screened for concatemer inserts of desired size. Utilizing DNA cassette concatemerization, an approximately 1500 bp multimer was isolated encoding a highly repetitive elastin-mimetic polypeptide (Figures 2 and 3).

The seamless cloning strategy was also used to generate a diblock (**AB**-type) copolymer consisting of the hydrophilic, elastin (**S1**) block at the N-terminus and the newly synthesized,

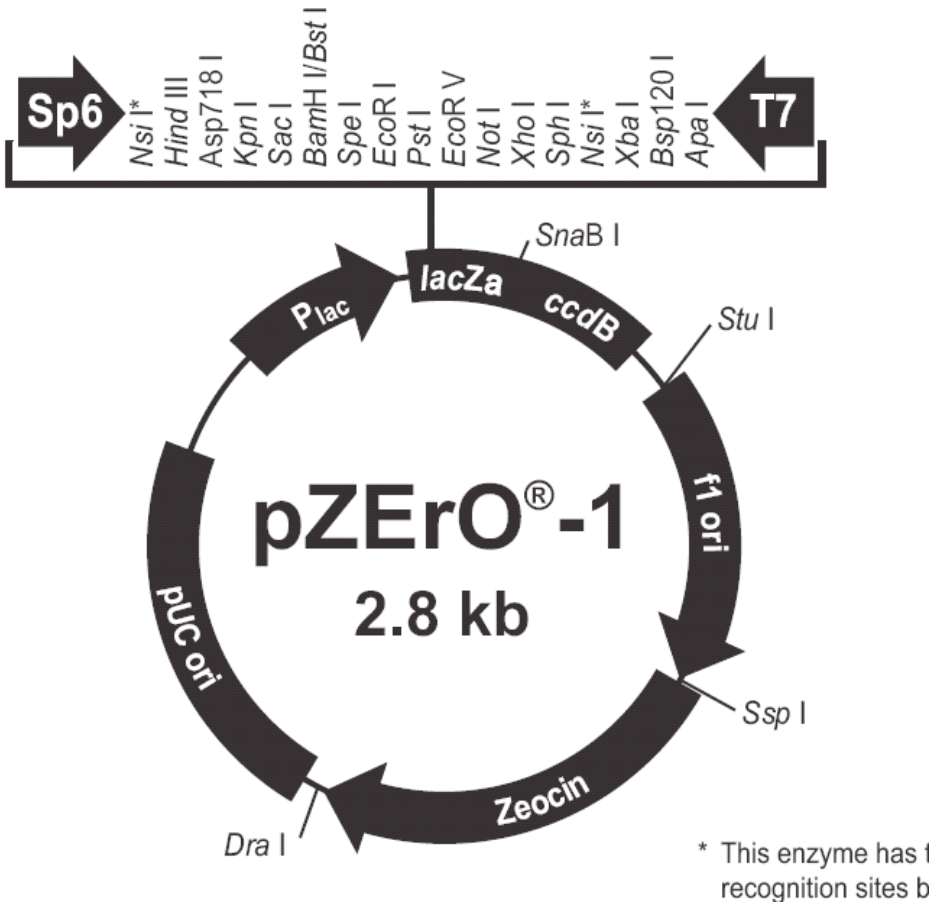


Figure 1. Plasmid map of the pZErO®-1 vector, from Invitrogen, Inc., indicating restriction endonuclease cleavage sites within the multiple cloning site. The restriction endonuclease recognition sites within the multiple cloning site are shown in detail. The elastin-M DNA monomer was cloned into the plasmid using the *Bam*H I/*Hind* III sites.

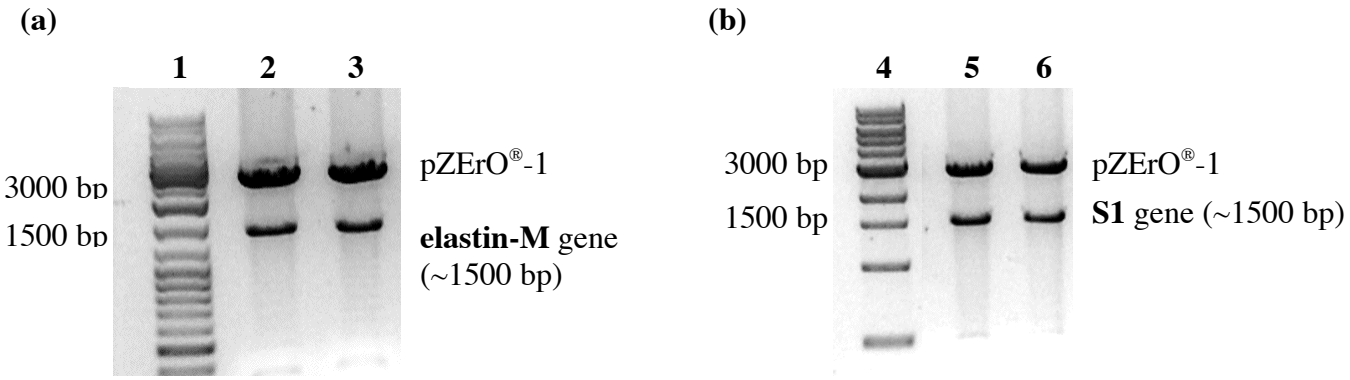


Figure 2. A 1% agarose gel depicting plasmids (a) **pMA20** and (b) **pMAP19** (~500 ng each) double digested with *Bam*H I and *Hin*d III, indicating the concatemer inserts. Twenty microliters of each digestion reaction were run on the gel in separate lanes. (a) Lane 1: O'GeneRuler™ 1 kb DNA ladder (Fermentas Inc.; Glen Burnie, MD) and lanes 2,3: The DNA from 2 single colonies transformed with correctly sequenced plasmid **pMAP20**, which contains the pZErO®-1 plasmid and elastin-M multimer insert. (b) Lane 4: O'GeneRuler™ 1 kb DNA ladder and lanes 5,6: The DNA from 2 single colonies transformed with correctly sequenced plasmid **pMAP19**, which contains the **S1** concatemer in the pZErO®-1 plasmid.

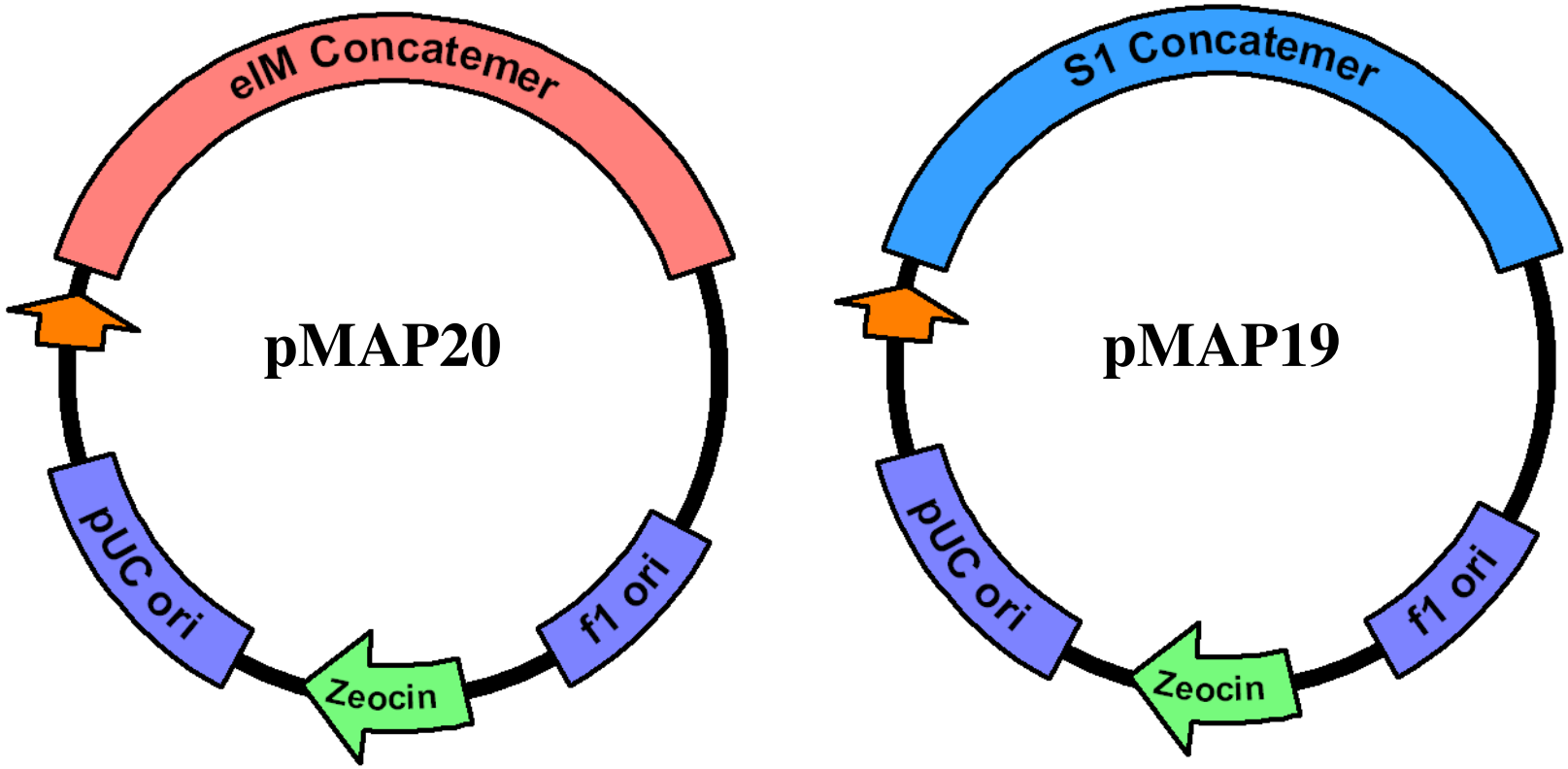


Figure 3. Plasmid **pMAP20** represents the recombinant pZErO®-1 plasmids following recloning of the pool of elastin-M concatemers in desired size range of 1000 to 2500 base pairs. Plasmid **pMAP19** represents cloning of the **S1** concatemer from the original recombinant plasmid into the pZErO®-1 plasmid between the *Bam*H I and *Hin*d III sites.

elastin-M concatemer at the C-terminus. The plasmids utilized and constructed in the synthesis of the **elastin-M** and **elastin-M** (**AB**) polymers are described in Table 1. The previously reported strategy[49] for directional assembly of a multiblock copolymer involves cleavage of the DNA concatemer cloning plasmids with a type IIs restriction endonucleases and an enzyme which cleaves within the antibiotic resistance gene of the cloning plasmid. The DNA fragments are purified by agarose gel electrophoresis and joined together via enzymatic ligation to produce a genetic fusion of the proximal and distal concatemer. Digestion within the antibiotic resistance gene reduces the incidence of false positives during subsequent selection of recombinant clones since ligation of the fragments reconstitutes the antibiotic marker. DNA concatemerization of the previously synthesized **S1** and elastin-M genes was conducted in cloning plasmids containing different antibiotic resistance genes. Thus, for efficient cloning of the diblock gene fusion, the **S1** concatemer was removed from the pZErO®-2.1 vector via digestion with *Hin*d III and *Bam*H I. The multimer was then purified via agarose gel electrophoresis and inserted into the corresponding *Hin*d III and *Bam*H I sites in the pZErO®-1 plasmid. Recombinant plasmids were screened for presence of the **S1** concatemer insert via agarose gel electrophoresis (Figure 2). The **pMAP19** and **pMAP20** cloning plasmids (Figure 3), which contain the **S1** and elastin-M concatemers, respectively, were utilized for directional assembly of a diblock (**AB**-type) plasmid. Recombinant plasmids were screened via agarose gel electrophoresis (Figure 4) and clones containing an insert of size concomitant with genetic fusion of the DNA concatemers were designated **pMAP23** (Figure 5).

The elastin-mimetic DNA cassettes were then cloned into a modified pQE-80L expression plasmid **pIL5** (Figures 6 and 7). The resulting plasmids **pMAP36** and **pMAP34**

Table 1. List of recombinant plasmids utilized and constructed in the synthesis of the **elastin-M** polymers.

Recombinant Plasmid Name	Plasmid/Insert	Comments
pSCP3.11	pZErO®-2.1/elastic (**S1**) concatemer	Zeo^R; Synthesis described in Chapter 2
pMAP18	pZErO®-1/**elastin-M** DNA monomer	Kn^R
pMAP19	pZErO®-1/elastin (**S1**) concatemer	
pMAP20	pZErO®-1/**elastin-M** concatemer	
pMAP23	pZErO®-1/**S1-elastin-M** diblock gene fusion	
pIL2	pQE-80L/**pQE Adaptor**	Amp^R
pIL5	pIL5 with removal of internal *Bsa* I site	
pMAP34	pIL5/**elastin-M** (**AB**)	Diblock gene fusion
pMAP36	pIL5/**elastin-M**	
pME1	pPROTetE.133/LarA.231	Cm^R, p15A ori; Reference[46]
pMetRS	pME1/*metG*	*metG* sequence Appendix 2
pHEC1	Modified pME1 lacking aaRS	Reference[51]

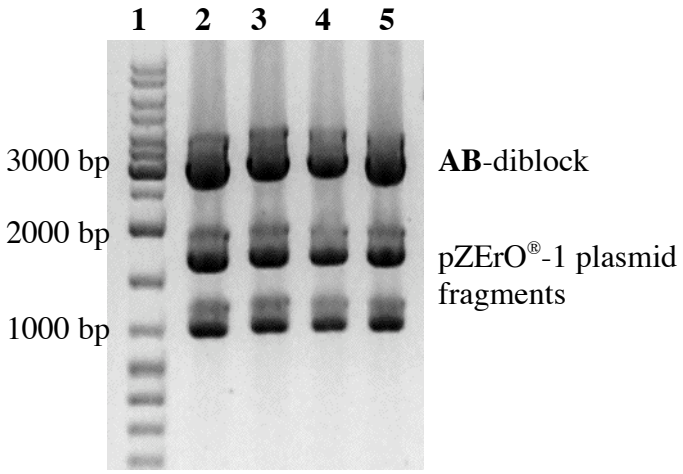


Figure 4. A 1% agarose gel depicting digestion with *Bbs* I, *Bam*H I, and *Dra* I of the **AB**-diblock formed by directional assembly of the **S1** and **elastin-M** concatemer cloning plasmids, **pMAP19** and **pMAP20**, respectively. Lane 1:O'GeneRuler™ 1 kb DNA ladder (Fermentas); and lanes 2-5: The DNA from 4 single colonies containing **pMAP23**.

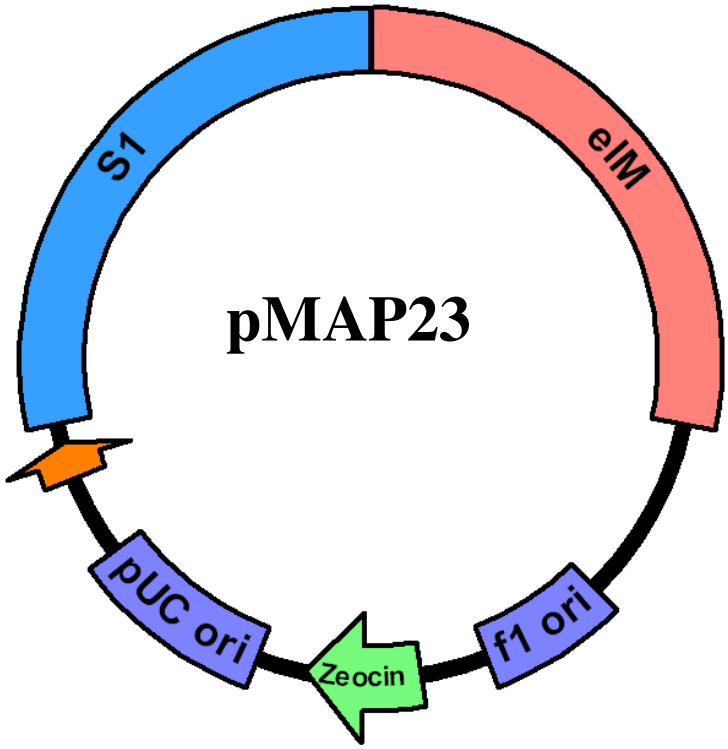


Figure 5. Plasmid **pMAP23** represents the recombinant pZErO®-1 plasmids containing the diblock (**AB**-type) gene fusion derived from elastin-mimetic sequences, where the **S1** concatemer is positioned at the N-terminus of the coding sequence and the elastin-M concatemer at the C-terminus.

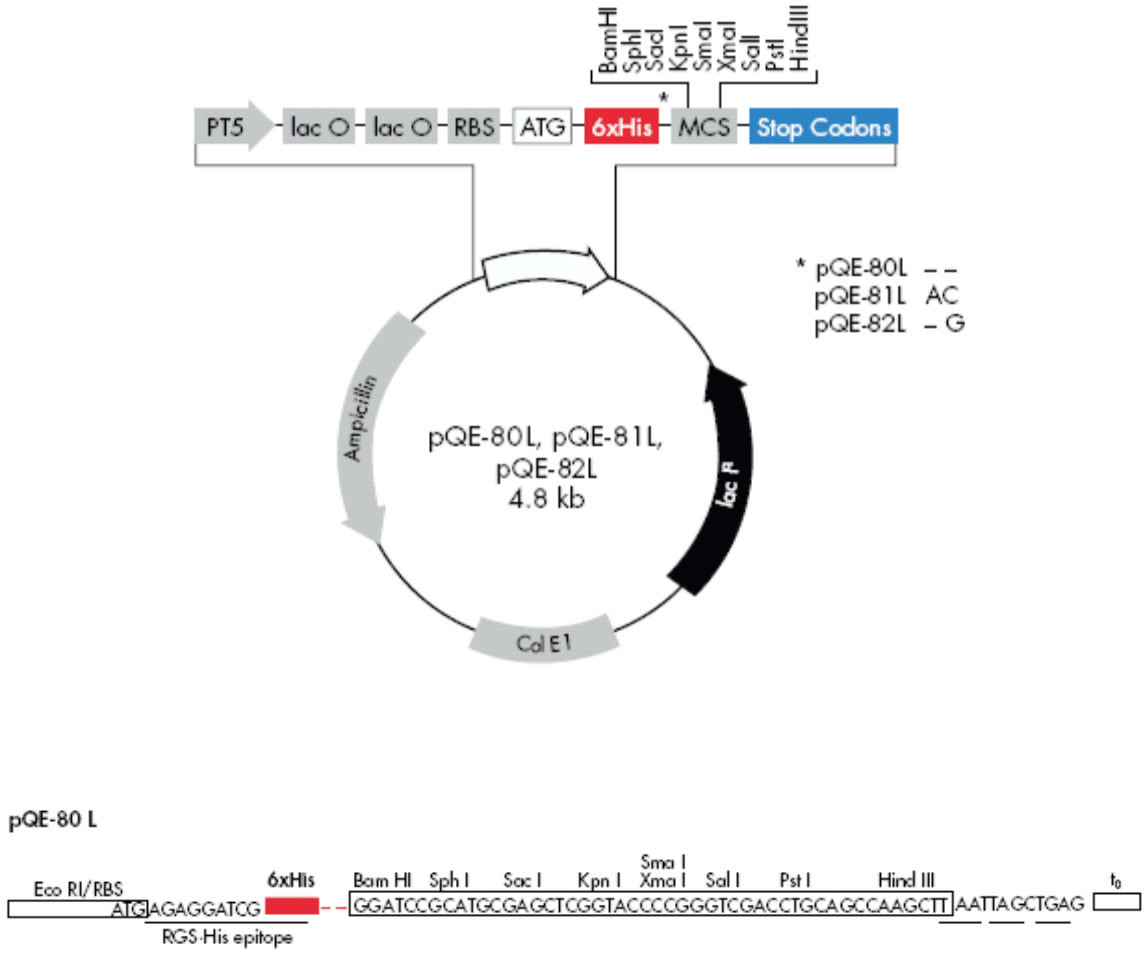


Figure 6. (**Above)** Plasmid map of the pQE-80L vector, from QIAGEN, indicating restriction endonuclease cleavage sites within the multiple cloning site. (**Below**) Expanded detail of the multiple cloning site (MCS).

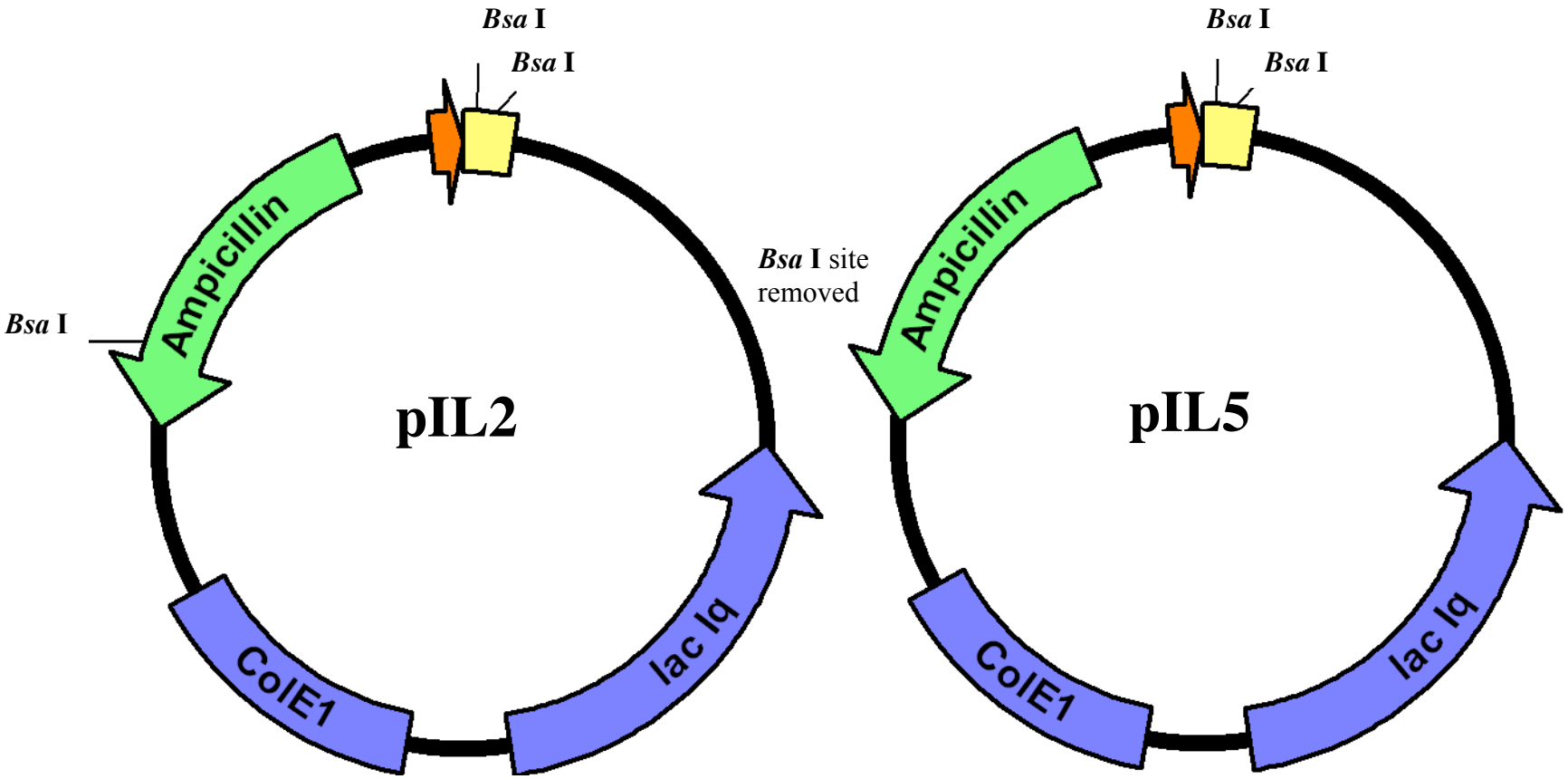


Figure 7. Plasmid **pIL2** was generated by cloning of the **pQE Adaptor** (shown in yellow) into the *Eco*R I and *Hin*d III sites within the multiple cloning site of pQE-80L. In order to facilitate cloning of the elastin-M concatemer cassettes into the modified expression plasmid, it was necessary to remove a *Bsa* I site internal to the pQE-80L vector sequence of **pIL2**. The modified plasmid was designated **pIL5** and utilized for subsequent cloning of the elastin-M cassettes.

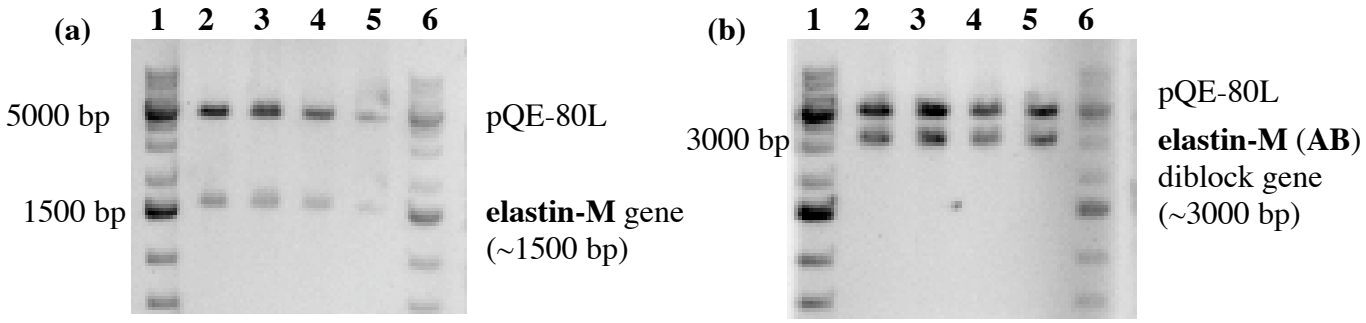


Figure 8. A 1% agarose gel demonstrating *Eco*R I and *Hin*d III double digestion of the plasmids (a) **pMAP36** and (b) **pMAP34** (~500 ng each) containing the elastin-M and elastin-M (AB) inserts, respectively, in the modified pQE-80L plasmid. Twenty microliters of each digestion reaction were run on the gel in separate lanes. (a) Lanes 1,6: O'GeneRuler^{TM} 1 kb Plus DNA ladder (Fermentas); and lanes 2-5: The DNA from 4 single colonies containing **pMAP36** plasmid. (b) Lanes 1,6: O'GeneRuler^{TM} 1 kb Plus DNA ladder; and lanes 2-5: The DNA from 4 single colonies containing the **pMAP34** plasmid.

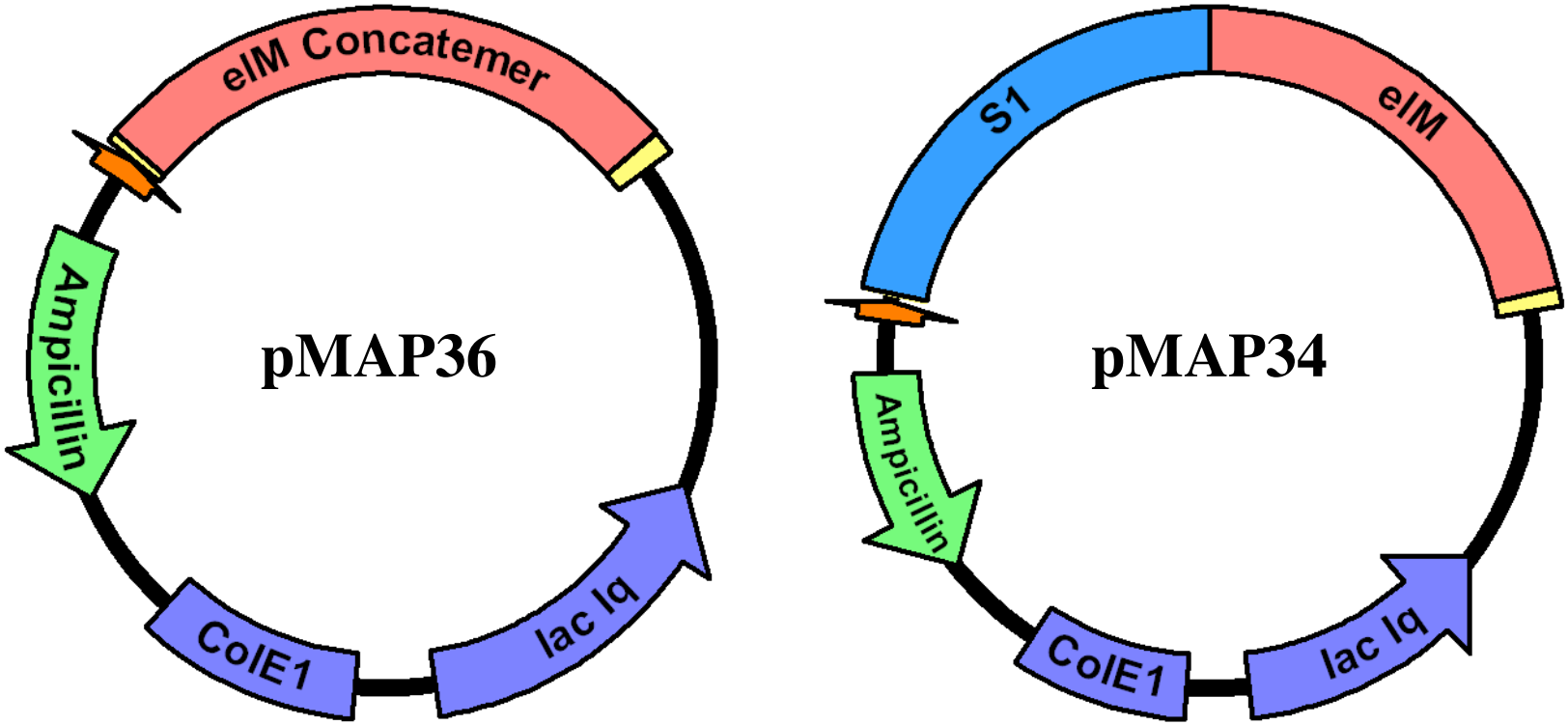


Figure 9. Plasmids **pMAP36** and **pMAP34** depict the modified pQE-80L expression plasmids that encode the **elastin-M** polymer and **elastin-M** (**AB**) diblock copolymer, respectively.

(Figures 8 and 9) place the elastin-mimetic polypeptides under control of the phage T5 promoter that is inducible with IPTG. In addition, the adaptor sequence was designed to fuse a decahistidine tag at the C-terminus of the elastin-mimetic polypeptide genes following cloning into the modified pQE-80L plasmid. The decahistidine tag could be useful for purification of the polymers via metal-affinity chromatography. In this study, however, purification of the elastin-mimetic polypeptides was accomplished by repetitive cycling through the inverse phase transition.

The **pMAP36** and **pMAP34** expression plasmids were utilized for directing synthesis of two elastin-mimetic proteins, **elastin-M** and **elastin-M** (**AB**), the latter of which corresponds to the diblock copolymer from genetic fusion of the **S1** and elastin-M concatemers. The elastin-mimetic polypeptides were targeted for global incorporation of the unnatural amino acid analogue L-**Hag** in response to methionine codons. The structurally similar analogue is incorporated into proteins in place of methionine under selective pressure during which the auxotrophic host strain is depleted of the natural amino acid but supplemented with the non-natural amino acid analogue[27]. Incorporation of the unnatural amino acid at multiple sites within the elastin-mimetic polypeptide was investigated using the auxotrophic host strain CAG18491 (*metE0-3079*::Tn10, Tc^R) under conditions of methionine depletion and analogue supplementation. Evidence of protein expression under these conditions supported incorporation of Hag at a level of efficiency approaching that of the canonical amino acid (Figure 11) for both the monoblock and diblock copolymers. Consistent with previous studies of elastin-mimetic polypeptides, the molar masses of the elastin-mimetic polymers as observed by SDS PAGE appear to be ~20% larger than the calculated molar masses[58,59].

Previous results suggest that increased MetRS activity may improve the overall yield of the analogue containing polymer during *in vivo* protein synthesis[28]. To investigate the role of increased wild-type aminoacyl-tRNA synthetase activity on the expression of the analogue containing elastin-mimetic polypeptides, an additional plasmid **pMetRS** (Figure 10) was constructed. In this vector, expression of the *E. coli* methionyl-tRNA synthetase is under control of an orthogonal promoter system. The vector **pME1** (Figure 10) was originally derived from the modular series of bacterial plasmids reported initially by Lutz and Bujard[60]. This plasmid has been used extensively in our lab for coexpression of wild type and mutant aminoacyl-tRNA synthetases as well as orthogonal tRNA/synthetase pairs[46,51,61] for amino acid analogue incorporation studies. The plasmid comprises the transcriptional/translational control elements, multiple cloning site, and chloramphenicol resistance marker of plasmid pPROTetE.133 and the compatible p15A replicon of plasmid pPROLarA.231. A duplex DNA cassette encoding the wild-type methionyl-tRNA synthetase (MetRS) was generated via PCR amplification of the metG gene from the *E. coli* genomic DNA. The amplification primers introduced *Kpn* I and *Xba* I cleavage sites at the 5' and 3' end of the PCR product, which was ligated into the corresponding sites within the polylinker of **pME1**. The MetRS cassette was placed under control of the P_L*tet* promoter and was expressed constitutively from the respective plasmid **pMetRS** in the *E. coli* auxotrophic strain CAG18491, which lacks the *tet* repressor. Electrophoretic analysis of whole cell lysates derived from small scale expression cultures indicated the methionine analogue displayed a detectable level of target protein expression with respect to the negative control cultures lacking methionine (Figure 11). These results indicate that coexpression of the wild-type aminoacyl-tRNA synthetase was not necessary for protein

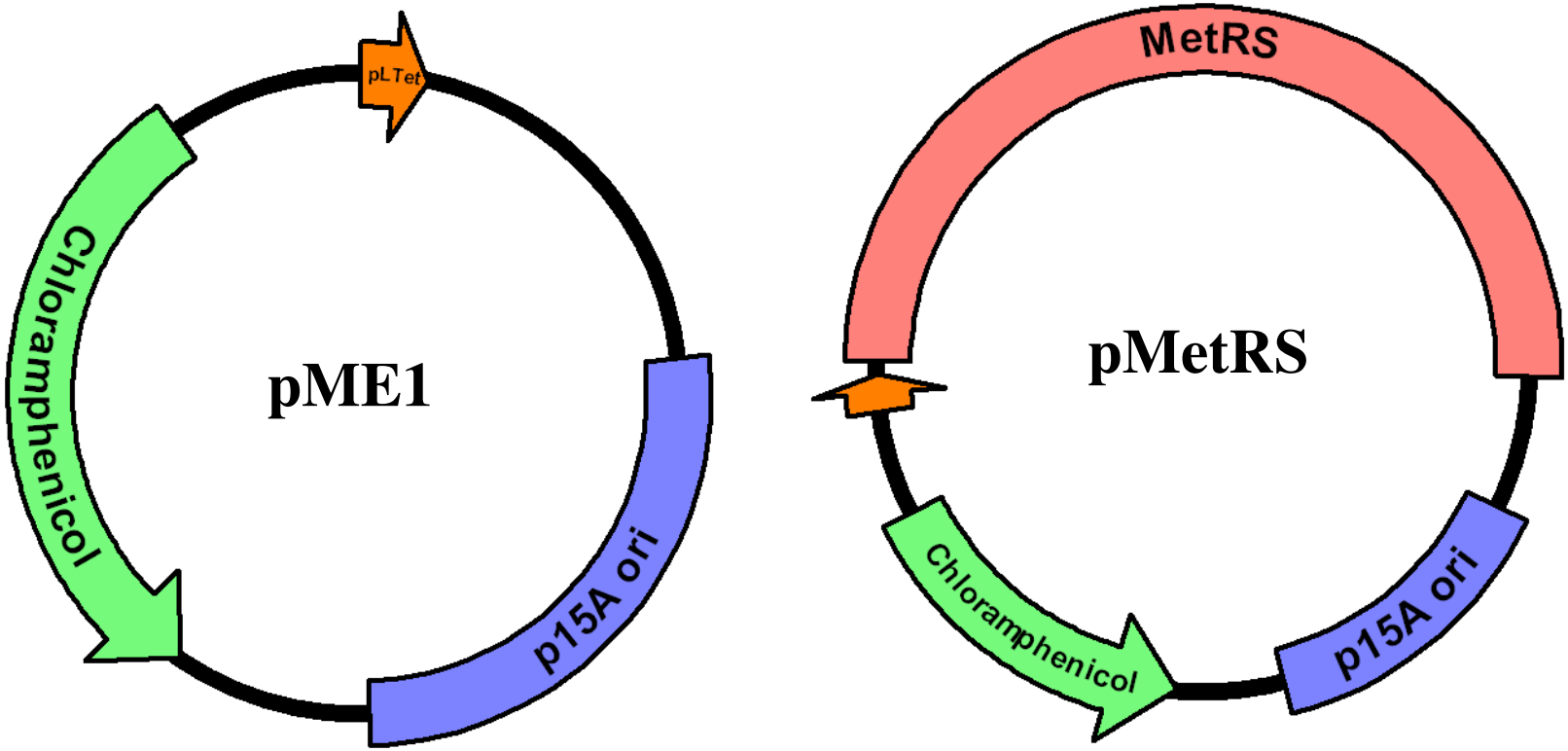


Figure 10. Plasmid **pME1** depicts the hybrid pPROTet/Lar plasmid previously synthesized[46] and was utilized here for expression of the wild type *Escherichia coli* methionyl-tRNA synthetase (MetRS). The plasmid **pMetRS** was generated by cloning of the PCR amplified *metG* gene into the *Eco*R I and *Hin*d III sites of the **pME1** plasmid. A modified **pME1** plasmid **pHEC1** (not pictured) was generated by inverse PCR amplification with primers designed to incorporate the *Nhe* I and *Pvu* I restriction endonuclease sites into the plasmid. The **pHEC1** plasmid was used for control experiments without coexpression of an additional copy of the methionyl-tRNA synthetase

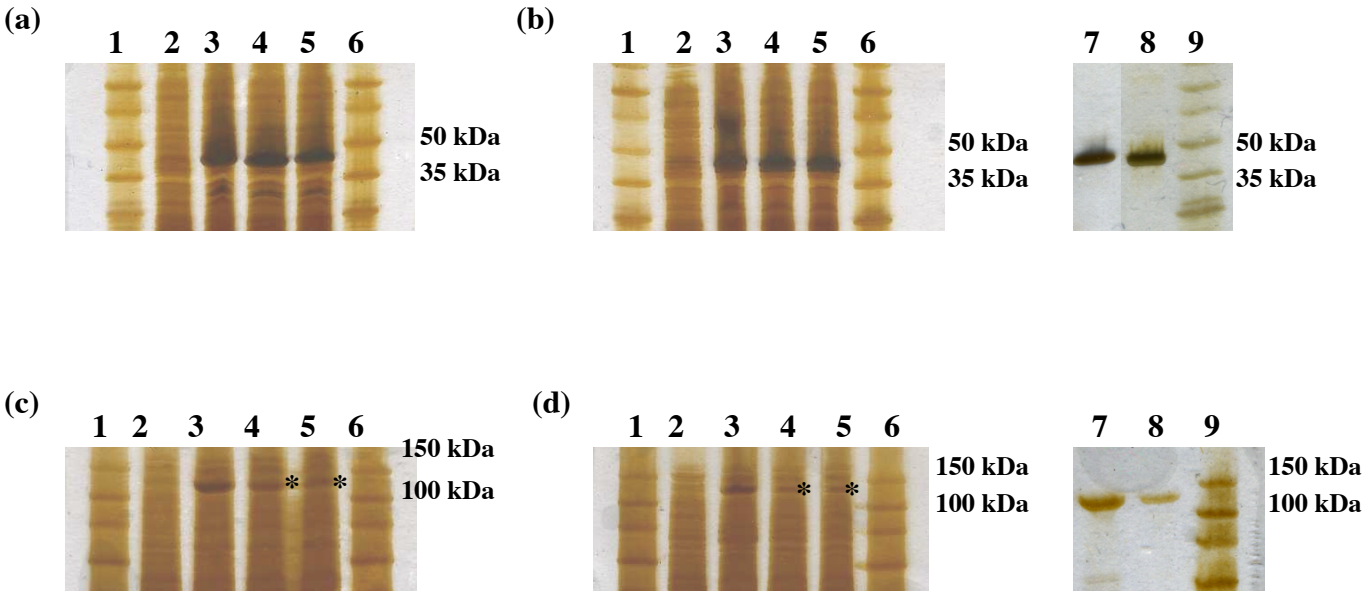


Figure 11. SDS PAGE with silver staining indicating multi-site incorporation of (*2S*)-2-amino-5-hexenoic acid (L-**Hag**) in the elastin-mimetic polymer (a, b) **elastin-M** and (c, d) **elastin-M** (**AB**) diblock copolymer. Elastin-mimetic polypeptides were synthesized from *E. coli* strain CAG18491 cotransformed with the plasmid encoding the artificial protein and a modified **pME1** plasmid as follows: (a) **pMAP36** and **pHEC1** (empty plasmid), (b) **pMAP36** and **pMetRS** (plasmid encoding additional copy of wild-type methionyl-tRNA synthetase, (c) **pMAP34** and **pHEC1**, and (d) **pMAP34** and **pMetRS**. Whole cell lysates from each expression culture were analyzed after 3 h induction with 1 mM IPTG. Lanes 1,6,9: Perfect Protein™ Marker, 15-150 kDa (Novagen®, a brand of EMD Chemicals Inc., Gibbstown, NJ); lane 2: methionine deficient (negative) control; lane 3: methionine supplemented (1 mM, positive) control; lane 4: L-**Hag** supplemented (0.5 mM); lane 5: L-**Hag** supplemented (1.0 mM); lane 7: purified protein from cultures supplemented with methionine (0.5 mM); lane 8: purified protein from cultures supplemented with L-**Hag** (0.5 mM). The expected molecular weights for the elastin-mimetic polypeptides **elastin-M** and **elastin-M** (**AB**) are approximately 40 kDa and 77 kDa, respectively. The **elastin-Hag** (**AB**) protein bands with supplementation of the non-natural amino acid analogue are indicated by asterisks (*).

production from cultures supplemented with the analogue L-**Hag**.

The elastin-derivates **elastin-M** and **elastin-M** (**AB**) incorporating the natural amino acid and the analogue were isolated from large scale cultures (1 L) using the inverse temperature cycling protocol[56]. The isolated yields of the polypeptides (Table 2) ranged from approximately 13 mg/L to 32 mg/L for fully induced expression cultures in minimal media (NMM) supplemented with 0.5 mM of the appropriate amino acid under conventional batch fermentation conditions in shake flask culture. The protein yields are comparable to previous reports[28] and especially impressive given the density of methionine residues in the target polypeptide sequence (Scheme 2), which reflects a higher degree of biosynthetic substitution that previously reported for incorporation of Hag[26]. Furthermore, protein yields for the **elastin-Hag** polymer from the CAG18491 expression strain coexpressing an additional copy of the methionyl-tRNA synthetase were improved almost 2-fold over coexpression with the empty (control) plasmid (13 mg/L). This result supports the previous observation that simple overexpression of MetRS can increase protein yields from cultures supplemented with analogue[28].

The identities of the elastin-mimetic polymers were confirmed via MALDI-TOF mass spectrometry (Table 2; Figures 12 and 13). The lower *m/z* ratio observed for the molecular ion in the mass spectra of **elastin-Hag** polymers, isolated from cultures supplemented with the analogue, indicate that the difference in mass is due to complete replacement of methionine by Hag. The molecular ions in the mass spectra are in agreement with calculated *m/z* values for each elastin-mimetic polymer, assuming the species are singly charged. The calculated *m/z* ratios for the polypeptides under investigation include retention of the N-terminal residue in the purified Met and Hag substituted polymers. The conditional *in vivo* excision of the initiator

Scheme 2. Amino Acid Sequence of Protein-Based Block Copolymers **elastin-M** and **elastin-M** (**AB**)

elastin-M M-[X]-VPGVGGSD$_4$KGH$_{10}$
elastin-M (AB) M[(VPGAG)$_2$VPGEG(VPGAG)$_2$]$_{19}$-[X]-VPGVGGSD$_4$KGH$_{10}$,

where [X] = [(VPGVG)$_2$VPGMG(VPGVG)$_2$]$_{18}$

Table 2. Expression yields and MALDI-TOF mass spectrometric data for elastin-mimetic polypeptides from large scale expression (1 L) in host strain CAG18491 cotransformed with a plasmid encoding the target elastin-mimetic polypeptides and the plasmid **pMetRS** with supplementation of 0.5 mM amino acid as indicated.

Elastin-mimetic Polypeptide	Yield (mg/L)	Calculated *m/z*[a]	Observed *m/z*	|*m/z*| (% error)
elastin-Met	32.0	40 150.60	40 158.25	7.65 (0.02)
elastin-Hag	25.4	39 769.65	39 771.20	1.55 (0.003)
elastin-Met (AB)	44.0	77 489.30	77 436.73	52.57 (0.07)
elastin-Hag (AB)	20.0	77 108.35	77 026.10	82.25 (0.10)

[a] Molar masses for the elastin-mimetic polypeptide **elastin-Hag** and **elastin-Hag** (**AB**) were calculated based on complete substitution of methionine with the amino acid analogue **Hag**.

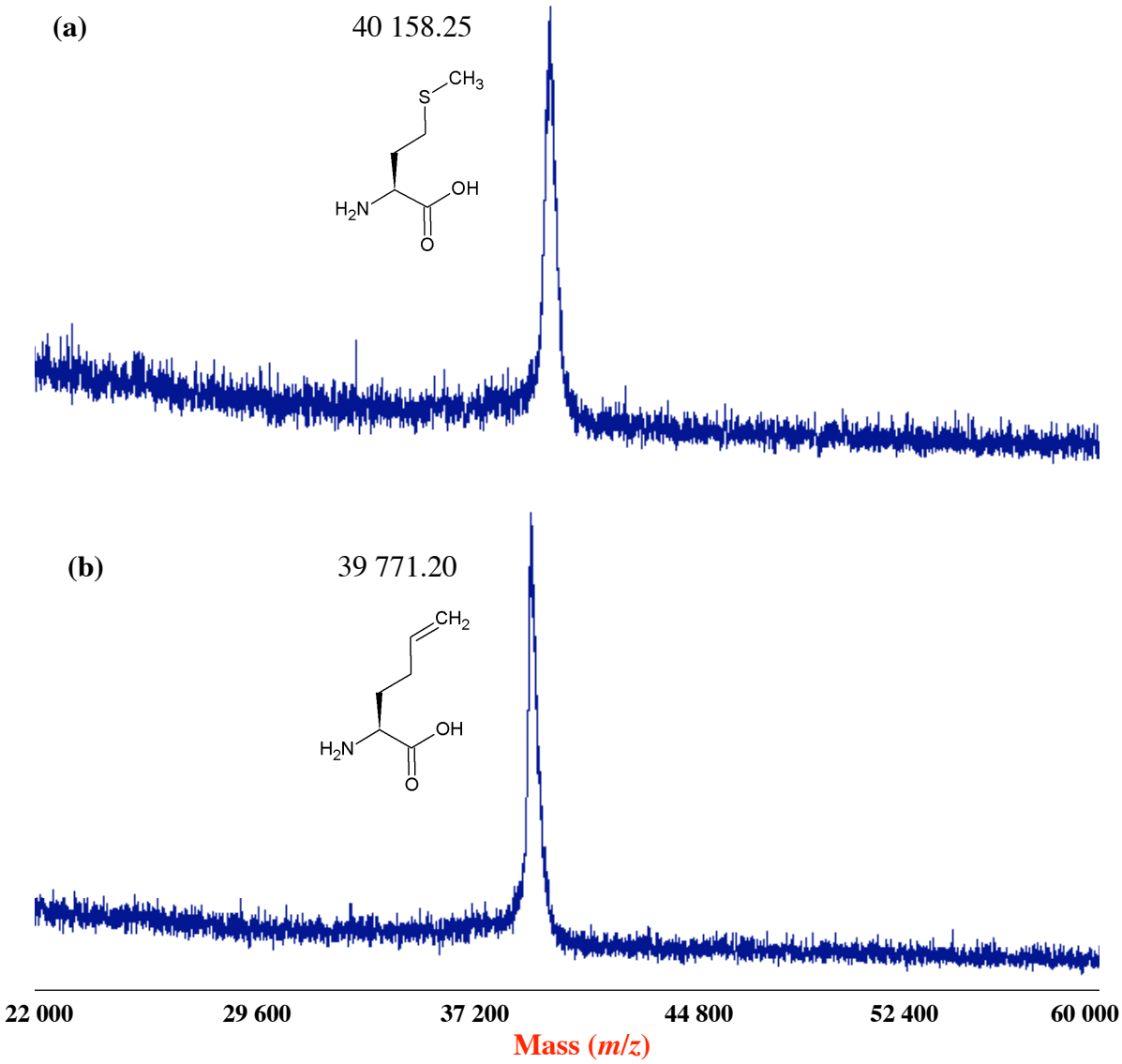


Figure 12. MALDI-TOF mass spectrometry of the **elastin-Met** and **elastin-Hag** polymers isolated from CAG18491 cells coexpressing plasmids **pMAP36** and **pMetRS** and supplemented with (a) L-methionine or (b) the analogue L-**Hag**. The peak in each spectrum is labeled and the chemical structure of the corresponding amino acid is shown. The lower *m/z* ratio observed for the molecular ion in the spectra of **elastin-Hag** is a result of replacement of methionine with the analogue and this difference (387.05) indicates complete substitution by the analogue.

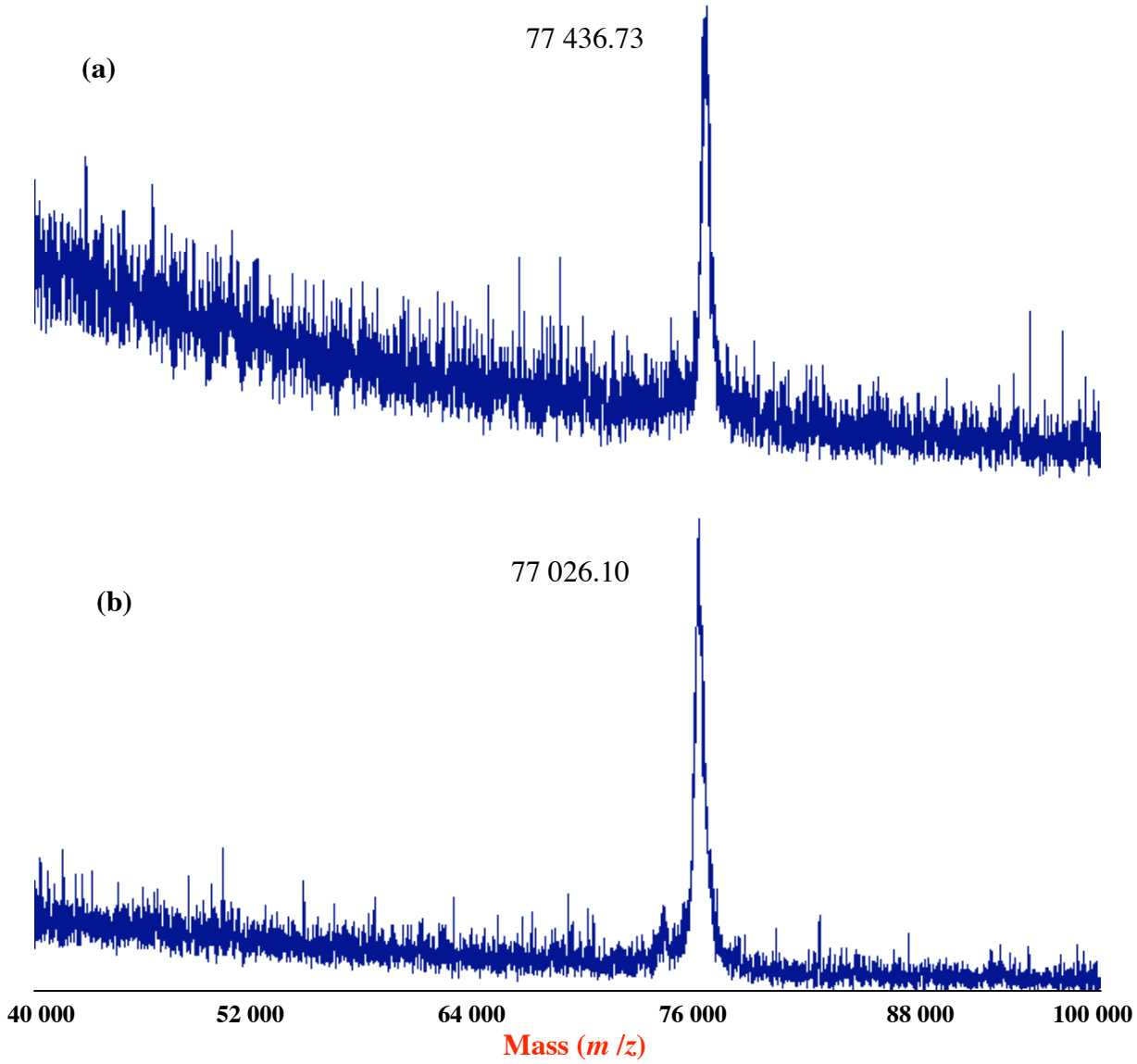


Figure 13. MALDI-TOF mass spectrometry of the **elastin-Met** (**AB**) and **elastin-Hag** (**AB**) diblock copolymers isolated from CAG18491 cells, with coexpression of plasmids **pMAP34** and **pMetRS**, supplemented with (a) L-methionine or (b) the analogue L-**Hag**. The peak in each spectrum is labeled. The lower *m/z* ratio observed for the molecular ion in the spectrum of **elastin-Hag** (**AB**) is a result of replacement of methionine with the analogue.

The conditional *in vivo* excision of the initiator methionine residue from *E. coli* proteins is correlated with the identity of the penultimate amino acid residue and the presence of a proline in the third position[62]. Specifically, in cases where the penultimate amino acid corresponds to valine, the removal of the N-terminal methionine is greatly reduced in model polypeptides. The molecular ions in the collected mass spectra for the elastin-mimetic polypeptides correspond well with calculated masses for elastin derivatives in which the encoded methionine residues were globally substituted with the analogue Hag, including at the N-terminus (Table 2).

Amino acid analysis of the purified **elastin-Hag** and **elastin-Hag** (**AB**) indicated that the proteins contained 0.1 mol% and 0.07 mol% methionine, respectively (Table 3). These values indicate a significant decrease in the methionine content compared to the **elastin-Met** and **elastin-Met** (**AB**) polymers, which contain 3.9 mol% and 1.8 mol% Met, respectively. The unnatural amino acid analogue is not detectable directly by amino acid analysis due to its instability under the analysis conditions. Assuming that the diminution in methionine content is due to replacement by Hag, the percent substitution by the analogue in the elastin-mimetic polypeptides may be estimated as a percentage of missing methionine. These calculations are consistent with high level substitution of the unnatural amino acid within the **elastin-Hag** and **elastin-Hag** (**AB**) polymers at 98.0% and 96.6 ± 5%, respectively.

Further evidence for incorporation of the alkene functionality into the polypeptide architecture was obtained via ^{1}H-NMR spectroscopy (Figure 14). Spectroscopic analysis of the **elastin-Hag** polymer demonstrated structural features that were commensurate with high levels of incorporation of the unnatural amino acid analogue, while these features are absent from spectra of the corresponding elastin-mimetic polypeptide containing the canonical amino acid,

Table 3. Amino acid compositional analysis of the purified elastin-mimetic polypeptides.

Amino Acid	elastin-Met	elastin-Hag	elastin-Met (AB)	elastin-Hag (AB)
		Calculated (Observed) [mol-%]		
Alanine (A)			8.0 (8.6)	8.0 (8.5)
Asparagine (D)	0.8 (1.0)	0.8 (1.0)	0.4 (0.5)	0.4 (0.9)
Glutamic Acid (E)			2.0 (2.1)	2.0 (3.5)
Glycine (G)	38.8 (39.8)	38.8 (41.1)	39.4 (40.6)	39.4 (39.9)
Histidine (H)	2.1 (0.6)	2.1 (1.6)	1.1 (1.0)	1.1 (1.2)
Lysine (K)	0.2 (0.3)	0.2 (0.3)	0.1 (0.2)	0.1 (0.3)
Methionine (M)	4.0 (3.9)	4.0 (0.1)	2.0 (1.8)	2.0 (0.07)
Proline (P)	19.2 (20.0)	19.2 (20.5)	19.6 (21.6)	19.6 (21.2)
Serine (S)	0.2 (0.2)	0.2 (0.2)	0.1 (0.1)	0.1 (0.3)
Valine (V)	34.6 (33.7)	34.6 (34.8)	27.3 (27.0)	27.3 (26.5)

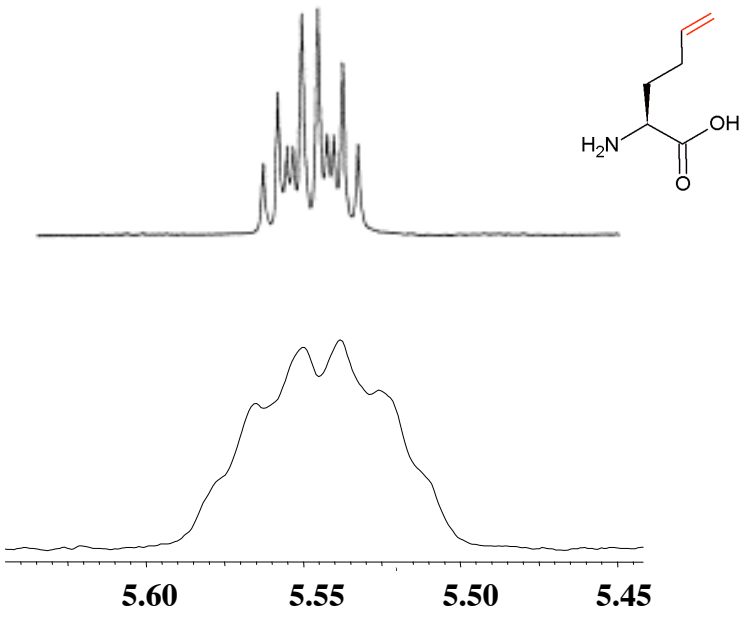


Figure 14. NMR spectroscopic analyses of the elastin derivatives. ^{1}H-NMR spectra of the unnatural amino acid L-**Hag** (upper spectrum[27]) and **elastin-Hag** (lower spectrum) in the downfield region encompassing the alkene proton resonance.

methionine. Specifically, comparison of the alkene regions of the ^{1}H-NMR spectra of the **elastin-Hag** polymer and the small molecule Hag show clearly the presence of a vinyl CH resonance of the unnatural amino acid side chain. These results support that L-**Hag** is a viable substrate during *in vivo* protein synthesis in an *E. coli* expression host and that the expression conditions are compatible with high levels of analogue substitution periodically within the highly repetitive elastin-mimetic polypeptide sequence.

Differential scanning calorimetry (DSC) provides a convenient method for determination of the thermodynamic properties associated with the elastin-mimetic polypeptide assembly from aqueous solution[63]. DSC measurements on dilute aqueous solutions (approximately 1 mg/mL) of the elastin-mimetic polypeptides **elastin-Met** and **elastin-Hag** displayed a sharp endothermic transition with a maximum between 31 °C and 32 °C (Figure 15), which was reversible upon cooling as demonstrated by rescan *in situ*. The temperature maximum indicated for the endothermic thermal transitions was consistent with the hydrophobicity scale for model elastin-mimetic polypeptides with substitution of methionine[64]. These results indicate the presence of the unnatural amino acid does not dramatically alter the temperature-dependent phase behavior of the **elastin-Met** polymer. The decrease in heat capacity of the polypeptides above the transition temperature (as indicated by the baseline drift, which is most pronounced for **elastin-Met** polymer solutions) is characteristic of the cooperative formation of condensed protein aggregates[65]. Finally, the temperatures of the thermal transitions did not significantly vary with changes in the pH of the copolymer solutions (Figures 16 and 17). This is expected for substitution at the fourth position by amino acid residues with nonpolar side chains. Finally, the van't Hoff enthalpies were calculated from the DSC data and are summarized in Table 4.

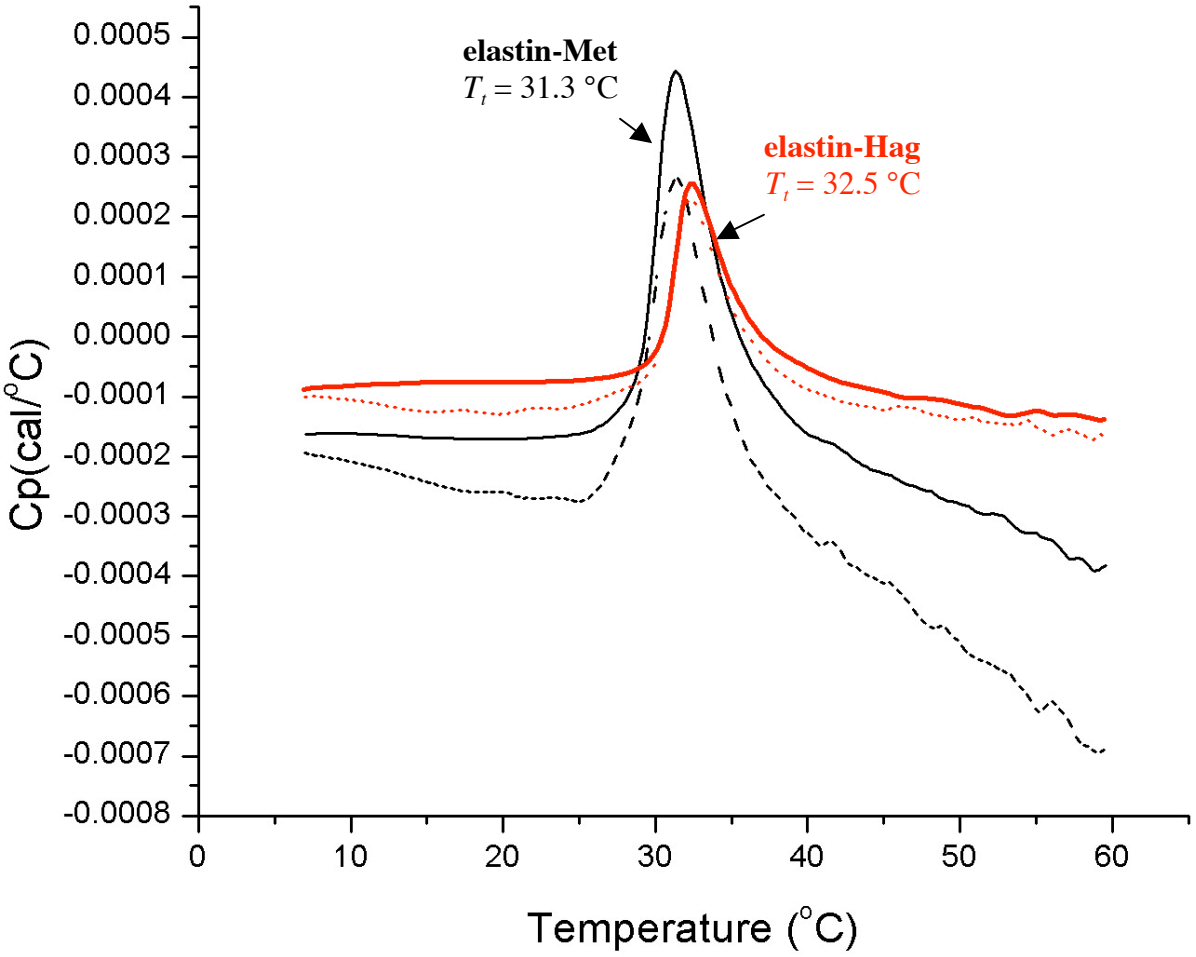


Figure 15. Raw DSC data for the endothermic thermal transitions of protein **elastin-Met** (shown in black) and **elastin-Hag** (shown in red) in dilute aqueous solutions. The temperature maximum for each transition is indicated. Rescan of the polymer solutions is shown by the dotted lines.

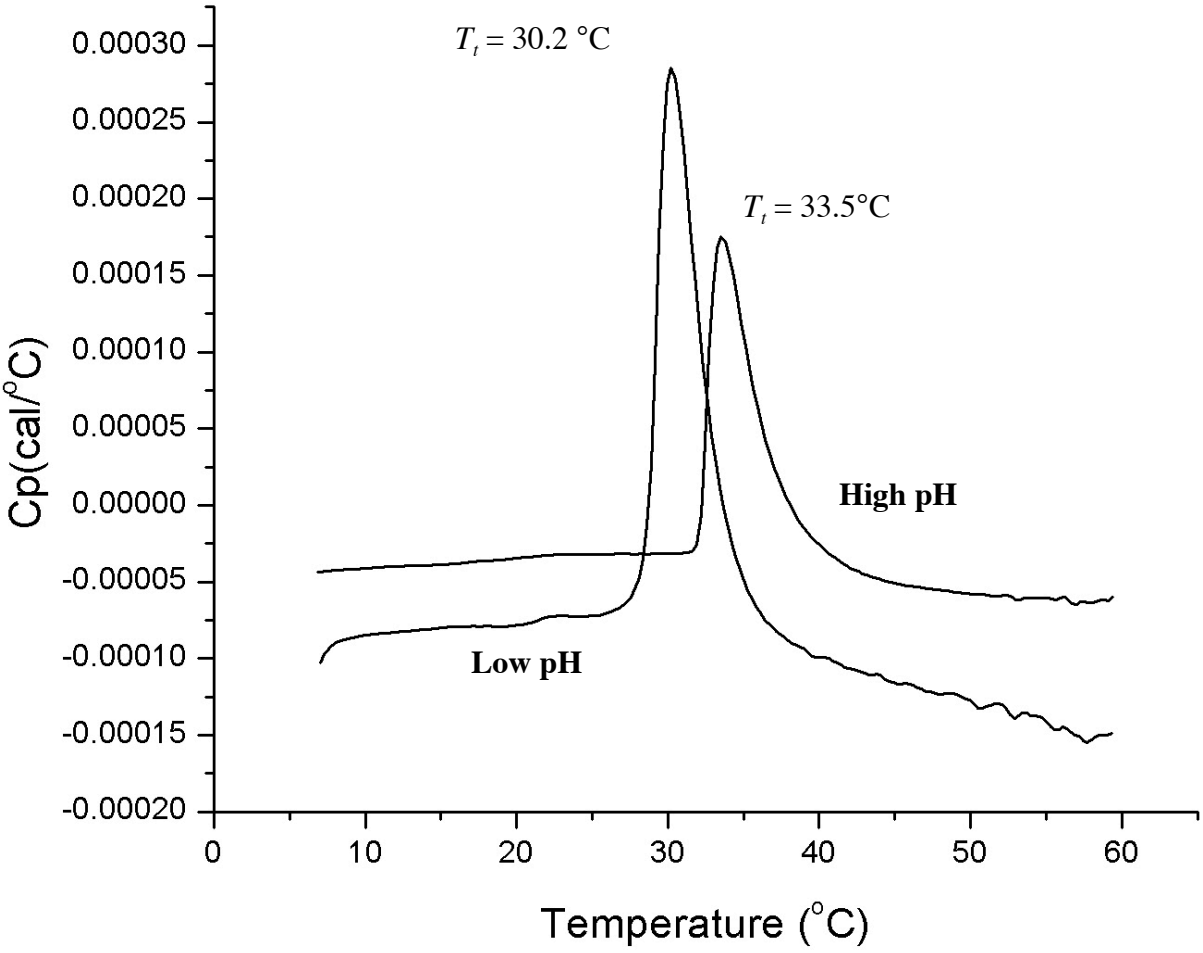


Figure 16. Raw DSC data for the endothermic thermal transition of protein **elastin-Met** in dilute aqueous solutions at high pH (40 mM NaOH) and low pH (40 mM acetic acid). The temperature maximum for each transition is indicated.

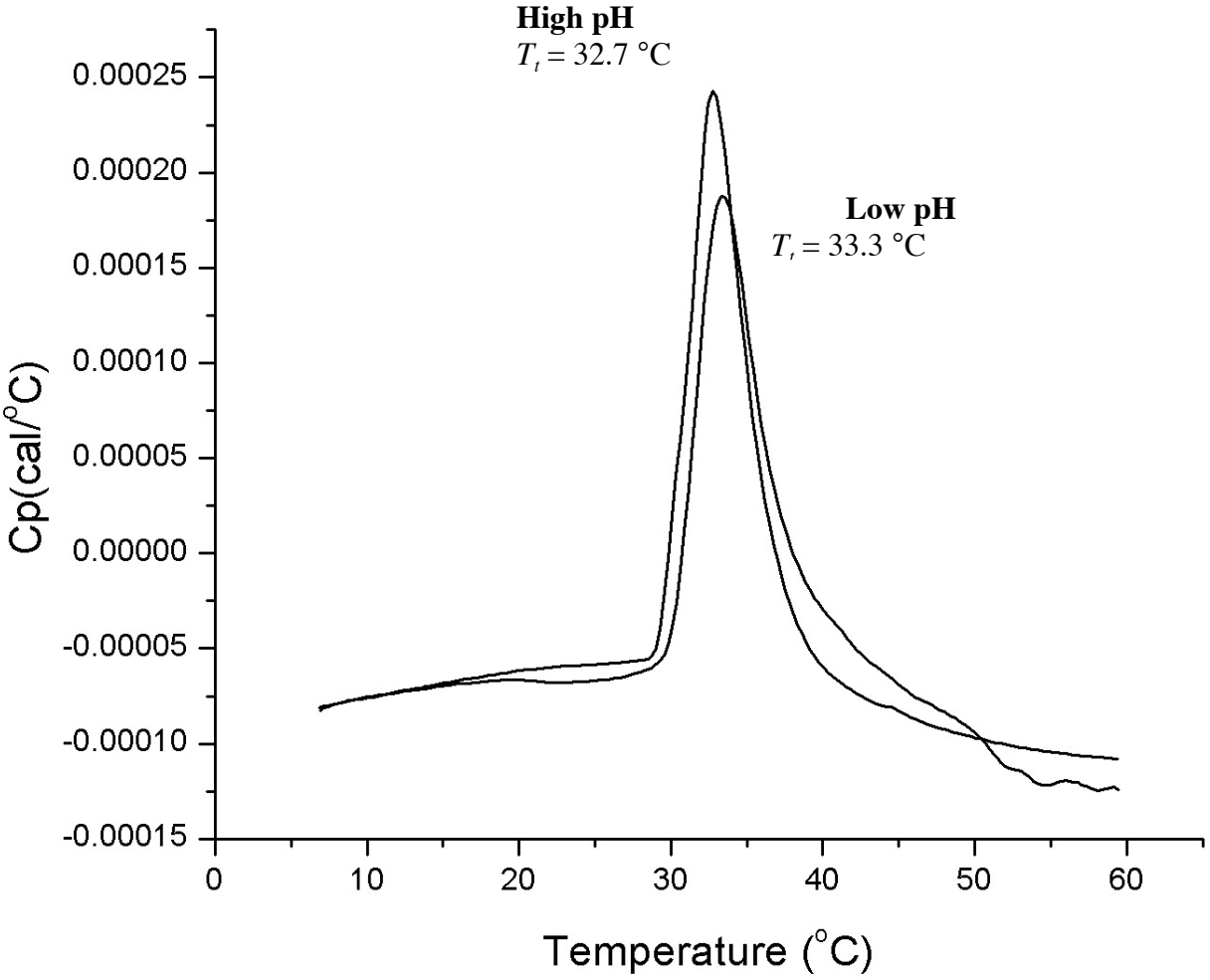


Figure 17. Raw DSC data for the endothermic thermal transition of protein **elastin-Hag** in dilute aqueous solutions at high pH (40 mM NaOH) and low pH (40 mM acetic acid). The temperature maximum for each transition is indicated.

Table 4. Cooperative van't Hoff transition enthalpies (ΔH) derived from raw differential scanning calorimetry (DSC) data for the **elastin-Met** and **elastin-Hag** polymer solutions.

	ΔH_{VH} (kcal mol^{-1})		
Protein	**Sterile ddH_2O**	**Low pH**	**High pH**
elastin-Met	213	153	106
elastin-Hag	150	134	148

Temperature turbidity experiments were conducted on solutions of the amphiphilic diblock copolymers substituted with methionine and the analogue **Hag** at low pH. The thermal transitions of the diblock copolymers were monitored by UV-visible spectrophotometry and evidenced by an increase in absorbance (Figure 18). These data indicated that temperature-driven self-assembly of the diblock copolymers occurred at temperatures higher than those reported for the monoblock polymers as examined by DSC. Specifically, assembly of the **elastin-Hag** (**AB**) copolymer occurred at a temperature approximately 7 °C to 10 °C higher than the transition observed for the corresponding monoblock **elastin-Hag** polymer. The temperature-dependent turbidimetry profile of the **elastin-Hag** (**AB**) copolymer suggests that assembly is initiated at approximately 37 °C followed by bulk aggregation of the copolymer over a narrow temperature range (approximately 48 °C to 52 °C). The shift in transition behavior of the diblock copolymers is most likely due to the presence of the glutamic acid residues in the hydrophilic domain. These results suggest that the formation of the coacervate and phase separation of the hydrophobic block is influenced by the solvent-polymer interactions of the hydrophilic domain. Furthermore, temperature turbidity profiles for the **elastin-Hag** (**AB**) copolymer in water indicate that aggregate formation is weak (Figure 18 inset) while temperature-dependent aggregation of aqueous solutions of the copolymer at high pH is undetectable below 85 °C. The temperature-dependent behavior of the **elastin-Hag** (**AB**) copolymer more closely mimics the temperature-dependent assembly expected for polymers comprising only the N-terminal hydrophilic block[64]. At high pH the transition may be undetectable due to the magnitude of the dominant attractive forces between the bound water and deprotonated glutamic acid residues of the hydrophilic domain.

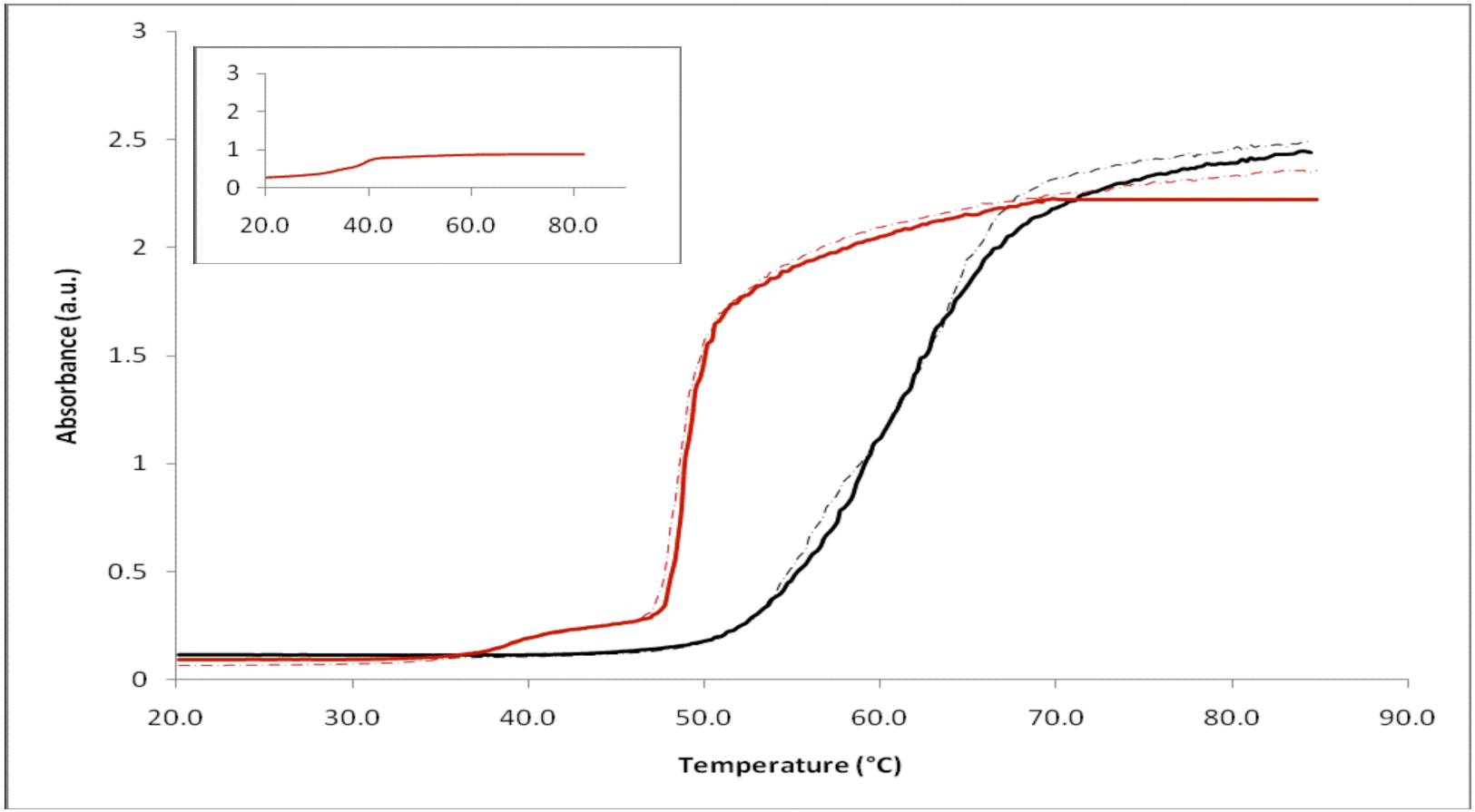


Figure 18. Temperature-dependent turbidimetry profiles for solutions of **elastin-Met** (**AB**) (shown in black) and **elastin-Hag** (**AB**) (shown in red) in dilute aqueous solutions at low pH (40 mM acetic acid). The reversibility of the transition is shown by rescan of the polymer solutions (dotted lines). The temperature-dependent turbidity profile for **elastin-Hag** (**AB**) in dilute aqueous solution at neutral pH is also shown (inset).

Previous studies in our lab have demonstrated that amphiphilic diblock (**AB**) copolymers derived from elastin-mimetic polypeptides can undergo selective segregation of the hydrophobic domain in aqueous solvents to form nanoscale micellar architectures, where the corona of the micelle is derived from the hydrophilic block (**A**) while the hydrophobic block (**B**) comprises the core[50,66]. The presence and morphology of micellar aggregates derived from self-assembly of **elastin-Hag** (**AB**) was investigated by transmission electron microscopy (TEM) of dilute aqueous solutions (~1 mg/mL) at both neutral pH and low pH (40 mM acetic acid). The copolymer solution was deposited onto the grid following formation of a coacervate at 37 °C, as indicated by the temperature-dependent turbidity profiles of diblock copolymer solutions. The most prominent morphology observed in the negatively stained TEM images was spherical nanoscale particles, characterized by relatively uniform diameters (Figure 19). Under acidic conditions additional morphologies such as high order globular and cylindrical aggregates of the spherical micelles were observed. It can be hypothesized that these higher order aggregates arise from further desolvation of the hydrophilic block and coalescence of the spherical nanoparticles. The occurrence of cylindrical micelles has also been observed in solutions of synthetic amphiphilic copolymers[67,68] as well as those derived from elastin-mimetic polypeptides[50]. The further aggregation of the micellar particles into an organized matrix, rather than a non-uniform dispersion over the grid surface, could potentially be enhanced by evaporative concentration during deposition of the copolymer solution.

(a) **(b)**

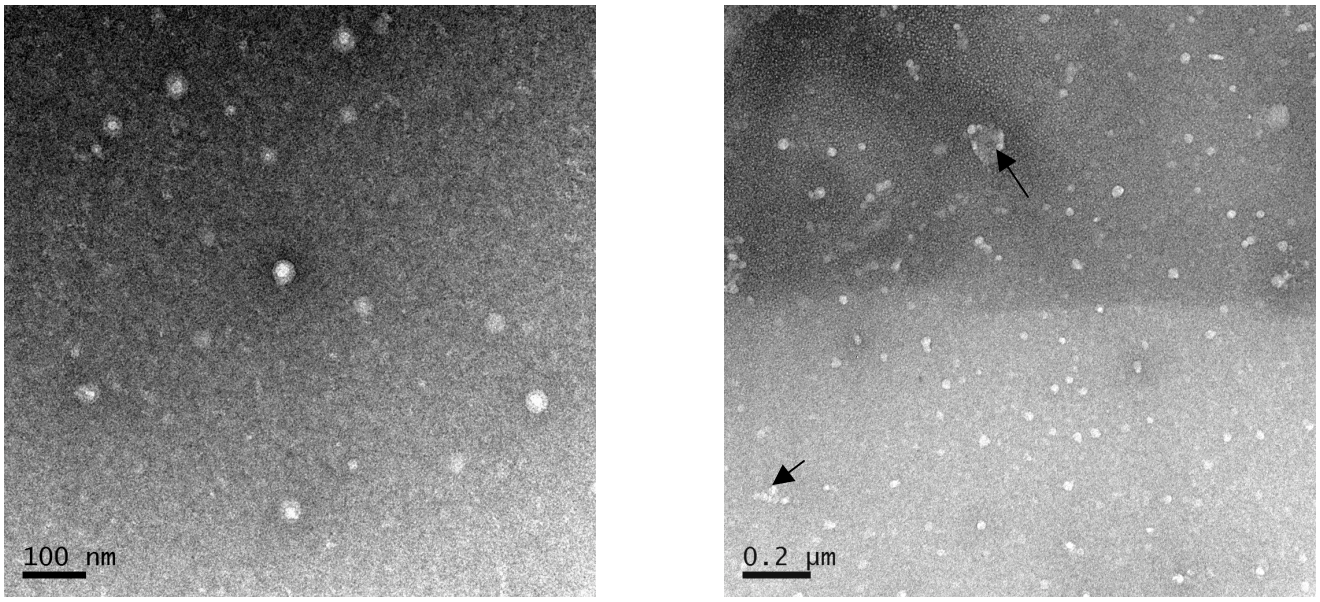


Figure 19. Representative negatively stained TEM images of the dispersed spherical nanoparticles derived from self-assembly of **elastin-Hag** (**AB**) in aqueous solution at (a) neutral pH and (b) pH 4. The spheres are characterized by uniform diameters. (b) At low pH, larger spherical aggregates and cylindrical particles (indicated by arrows) have arisen due to further desolvation of the hydrophilic block and subsequent coalescence of the spherical aggregates.

Conclusion

In summary, we have established an *E. coli* expression strain that is competent for incorporation of an unnatural amino acid with novel alkene functionality and developed culture conditions that result in virtually complete substitution of the analogue within an artificial protein. The method has been employed to generate novel elastin-mimetic polypeptides in which the analogue L-**Hag** serves as a viable substrate during *in vivo* protein synthesis. The residue-specific incorporation strategy permits global substitution of the analogue in response to the methionine codon at nineteen positions, representing an especially high degree of biosynthetic substitution. Both the monoblock and diblock proteins, under conditions favorable for hydrophobic aggregation, underwent a reversible sol-gel transition over a narrow temperature range with an inverse temperature profile characteristic of model elastin-mimetic polypeptides. Furthermore, the diblock copolymer was characterized by the formation of nanoscale micellar morphologies at a temperature above its T_t. We envision that these novel elastin-mimetic polypeptides containing the unnatural amino acid Hag will have advantages over previously designed polymers. The terminal vinyl groups present periodically within the protein-based polymer can serve as attractive sites for photo-initiated cross-linking under physiologically relevant conditions and further derivatization via the versatile chemistry of alkenes.

References

[1] Hecht, S. M. *Accounts of Chemical Research* **1992**, *25*, 545-552.

[2] Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S. *Journal of Biological Chemistry* **1978**, *253*, 4517-4520.

[3] Heckler, T. G.; Roesser, J. R.; Xu, C.; Chang, P. I.; Hecht, S. M. *Biochemistry* **1988**, *27*, 7254-7262.

[4] Bain, J. D.; Diala, E. S.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. *Journal of the American Chemical Society* **1989**, *111*, 8013-8014.

[5] Bain, J. D.; Switzer, C.; Chamberlin, R.; Bennert, S. A. *Nature* **1992**, *356*, 537-539.

[6] Cornish, V. W.; Mendel, D.; Schultz, P. G. *Angewandte Chemie International Edition* **1995**, *34*, 621-633.

[7] Gallivan, J. P.; Lester, H. A.; Dougherty, D. A. *Chemistry and Biology* **1997**, *4*, 739-749.

[8] Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M. *Journal of the American Chemical Society* **1999**, *121*, 34-40.

[9] Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182-188.

[10] Nowak, M. W.; Kearney, P. C.; Sampson, J. R.; Saks, M. E.; Labarca, C. G.; Silverman, S. K.; Zhong, W.; Thorson, J.; Abelson, J. N.; Davidson, N.; Schultz, P. G.; Dougherty, D. A.; Lester, H. A. *Science* **1995**, *268*, 439-442.

[11] Robertson, S. A.; Ellman, J. A.; Schultz, P. G. *Journal of the American Chemical Society* **1991**, *113*, 2722-2729.

[12] Saks, M. E.; Sampson, J. R.; Nowak, M. W.; Kearney, P. C.; Du, F.; Abelson, J. N.; Lester, H. A.; Dougherty, D. A. *Journal of Biological Chemistry* **1996**, *271*, 23169-23175.

[13] Turcatti, G.; Nemeth, K.; Edgerton, M. D.; Meseth, U.; Talabot, F.; Peitsch, M.; Knowles, J.; Vogel, H.; Chollet, A. *Journal of Biological Chemistry* **1996**, *271*, 19991-19998.

[14] Furter, R. *Protein Science* **1998**, *7*, 419-426.

[15] Ibba, M.; Hennecke, H. *FEBS Letters* **1995**, *364*, 272-275.

[16] Liu, D. R.; Magliery, T. J.; Pasternak, M.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 10092-10097.

[17] Liu, D. R.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 4780-4785.

[18] Cowie, D. B.; Cohen, G. N. *Biochimica et Biophysica Acta* **1957**, *26*, 252-261.

[19] Hortin, G.; Boime, I. *Methods in Enzymology* **1983**, *96*, 777-784.

[20] Richmond, M. H. *Bacteriology Reviews* **1962**, *26*, 398-420.

[21] Wilson, M. J.; Hatfield, D. L. *Biochimica et Biophysica Acta* **1984**, *781*, 205-215.

[22] *Comprehensive Organic Synthesis*; Pergamon Press: Oxford, 1991.

[23] Budisa, N.; Steipe, B.; Demange, P.; Eckerskorn, C.; Kellermann, J.; Huber, R. *European Journal of Biochemistry* **1995**, *230*, 788-796.

[24] Duewel, H.; Daub, E.; Robinson, V.; Honek, J. F. *Biochemistry* **1997**, *36*, 3404-3416.

[25] Jakubowski, H. *Journal of Biological Chemistry* **2000**, *275*, 21813-21816.

[26] Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Letters* **2001**, *502*, 25-30.

[27] van Hest, J. C. M.; Tirrell, D. A. *FEBS Letters* **1998**, *428*, 68-70.

[28] Kiick, K. L.; Tirrell, D. A. *Tetrahedron* **2000**, *56*, 9487-9483.

[29] Mechulam, Y.; Schmitt, E.; Maveyraud, L.; Zelwer, C.; Nureki, O.; Yokoyama, S.; Konno, M.; Blanquet, S. *Journal of Molecular Biology* **1999**, *294*, 1287-1297.

[30] Serre, L.; Verdon, G.; Choinowski, T.; Hervouet, N.; Risler, J.-L.; Zelwer, C. *Journal of Molecular Biology* **2001**, *306*, 863-876.

[31] Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. L.; Davis, B. G. *Journal of the American Chemical Society* **2008**, *130*, 9642-9643.

[32] Lin, Y. A.; Chalker, J. M.; Davis, B. G. *ChemBioChem* **2009**, *10*, 959-969.

[33] Biagini, S. C. G.; Gibson, S. E.; Keen, S. P. *journal of the Chemical Society, Perkin Transactions 1* **1998**, 2485-2500.

[34] Gibson, S. E.; Gibson, V. C.; Keen, S. P. *Chemical Communications* **1997**, 1107-1108.

[35] Discher, D. E.; Janmey, P.; Wang, Y.-l. *Science* **2005**, *310*, 1139-1143.

[36] Cukierman, E.; Pankov, R.; Stevens, D. R.; Yamada, K. M. *Science* **2001**, *294*, 1708-1712.

[37] Pelham, R. J., Jr.; Wang, Y.-l. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 13661-13665.

[38] Engler, A.; Bacakova, L.; Newman, C.; Hategan, A.; Griffin, M.; Discher, D. E. *Biophysical Journal* **2004**, *86*, 617-628.

[39] Engler, A. J.; Richert, L.; Wong, J. Y.; Picart, C.; Discher, D. E. *Surface Science* **2004**, *570*, 142-154.

[40] Gray, D. S.; Tien, J.; Chen, C. S. *Journal of Biomedical Materials Research Part A* **2003**, *66A*, 605-614.

[41] Lo, C.-M.; Wang, H.-B.; Dembo, M.; Wang, Y.-l. *Biophysical Journal* **2000**, *79*, 144-152.

[42] Wang, H.-B.; Dembo, M.; Wang, Y.-l. *American Journal of Physiology; Cell Physiology* **2000**, *279*, C1345-C1350.

[43] Engler, A. J.; Griffin, M. A.; Sen, S.; Bönnemann, C. G.; Sweeney, H. L.; Discher, D. E. *Journal of Cell Biology* **2004**, *166*, 877-887.

[44] Engler, A. J.; Rehfeldt, F.; Sen, S.; Discher, D. E. *Methods in Cell Biology* **2007**, *83*, 521-545.

[45] Engler, A. J.; Sen, S.; Sweeney, H. L.; Discher, D. E. *Cell* **2006**, *126*, 677-689.

[46] Kim, W.; George, A.; Evans, M.; Conticello, V. P. *ChemBioChem* **2004**, *5*, 928-936.

[47] Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. *Angewandte Chemie International Edition* **2000**, *39*, 2148-2152.

[48] van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. *Journal of the American Chemical Society* **2000**, *122*, 1282-1288.

[49] Payne, S. C.; Patterson, M.; Conticello, V. P. In *Protein Engineering Handbook*; Lutz, S., Bornscheuer, U. T., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2009; Vol. 1, p 915-936.

[50] Lee, T. A. T.; Cooper, A.; Apkarian, R. P.; Conticello, V. P. *Advanced Materials* **2000**, *12*, 1105-1110.

[51] Carpenter, H. E., Emory University, 2008.

[52] Nichols, B. P.; Shafiq, O.; Meiners, V. *Journal of Bacteriology* **1998**, *180*, 6408-6411.

[53] Singer, M.; Baker, T. A.; Schnitzler, G.; Deischel, S. M.; Goel, M.; Dove, W. F.; Jaacks, K.; Grossman, A. D.; Erickson, J.; Gross, C. A. *Microbiology Reviews* **1989**, *53*, 1-24.

[54] Sambrook, J.; Russell, D. W. *Molecular cloning: a laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

[55] Ehrt, S.; Schnappinger, D. *Methods in Molecular Biology* **2003**, *235*, 79-82.

[56] Meyer, D. E.; Chilkoti, A. *Nature Biotechnology* **1999**, *17*, 1112-1115.

[57] Wishart, D. S.; Bigam, C. G.; Yao, J.; Abilgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. *Journal of Biomolecular NMR* **1995**, *6*, 135-140.

[58] Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2002**, *3*, 357-367.

[59] Trabbic-Carlson, K.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2003**, *4*, 572-580.

[60] Lutz, R.; Bujard, H. *Nucleic Acids Research* **1997**, *25*, 1203-1210.

[61] Anderson, J. R., Emory University, 2004.

[62] Hirel, H.; Schmitter, J.-M.; Dessen, P.; Fayat, G.; Blanquet, S. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86*, 8247-8251.

[63] Luan, C.-H.; Harris, R. D.; Prasad, K. U.; Urry, D. W. *Biopolymers* **1990**, *29*, 1699-1706.

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

[65] Cooper, A. *Biophysical Chemistry* **2000**, *85*, 25-39.

[66] Wright, E. R.; Conticello, V. P. *Advanced Drug Delivery Reviews* **2002**, *54*, 1057-1073.

[67] Won, Y.-Y.; Davis, H. T.; Bates, F. S. *Science* **1999**, *283*, 960-963.

[68] Zhang, L.; Eisenberg, A. *Science* **1995**, *268*, 1728-1731.

Chapter 4

Cotranslational incorporation of *L*-DOPA into an elastin-mimetic polymer in an *E. coli* expression system

Introduction

The discovery of technologically useful hydrogels is of growing interest because of their potential application in the fields of bioengineering and biomaterials[1,2]. The development of polymeric hydrogel materials has the potential to impact the generation of synthetic analogues for tissue engineering and growth, drug delivery agents, and cell growth scaffolds. The use of protein-based polymers is an attractive strategy for the development of such materials with controlled macromolecular architecture, mechanical, chemical, and biological properties, and biocompatibility. The exquisite design features found in native proteins have influenced many efforts in this area. In our lab, we have worked extensively to design and synthesize a new class of polymers derived from the consensus pentapeptide repeat motif (Val-Pro-Gly-Xaa-Gly) of native elastin, which can form physically[3-5] and covalently[6,7] cross-linked hydrogel matrices. Strategies for covalently cross-linking elastin-like polypeptides have taken a cue from nature, where formation of the native elastin network occurs following cross-linking of lysine residues within the polypeptide chain of in the soluble precursor tropoelastin. Generally, elastin-mimetic polypeptides are cross-linked through lysine residues at position Xaa using chemical cross-linkers including isocyanates, NHS-ester, phosphines, aldehydes, and geninpin[5,6,8-18]. These strategies can be slow, inefficient, and limited by the cytotoxicity of the chemical agent. In an effort to overcome these limitations, chemists and biochemists have again looked to the strategies employed by nature to generate cross-linked hydrogels from native polymer substrates.

The unnatural amino acid L-3,4-dihydroxyphyenylalanine (**DOPA**) (Scheme 1) is found in a diverse number of species from bacteria to humans performing a wide array of functions[19,20]. DOPA is an important neurotransmitter produced naturally in humans by tyrosinase-catalyzed

oxidation of tyrosine in the biosynthesis of melanin and by tyrosine hydroxylase in the brain for catecholamine neurotransmitter biosynthesis. The unnatural amino acid also plays an important role in the adhesive and cross-linking properties of proteins generally found in marine invertebrates, where it is generated through post-translational modification of genetically encoded tyrosine residues[21-23]. The catecholic side chain (Scheme 1) provides access to chemistry inaccessible to the 20 genetically encoded amino acids. Specifically, deprotonation of the *ortho*-hydroxyl groups transforms DOPA into an avid metal chelator. Hence, catechol-containing molecules are commonly found in natural systems responsible for metal solubilization, transport, and storage. In addition, DOPA is able to undergo 1 or 2 electron oxidations to yield semiquinone radicals or quinone species[19,22,24]. These groups may react further in a variety of proposed bond forming mechanisms including Michael additions, Schiff base formation, and ring-ring coupling reactions[21,22,24] (Figure 1[21]). Through an intermolecular cross-linking reaction, the DOPA residues of mussel adhesive protein are joined together to form a solid adhesive plaque, which mediate the firm attachment of the organism to a wide variety of wet surfaces[25].

In an effort to mimic the mechanical properties of the native mussel adhesive protein network, a number of groups have targeted the generation of synthetic polypeptide analogues by incorporation of the unnatural amino acid into polymer backbones, side chains, and end groups[26-35]. In general, incorporation of DOPA into polypeptides has utilized *in vitro* protein synthesis and solid-phase peptide synthesis methods[36]. For example, Yamamoto and coworkers have prepared DOPA-containing polymers, including poly(DOPA) and polypeptides consisting of repetitive sequences derived from the adhesive domains of proteins found in marine organisms[30-33]. The

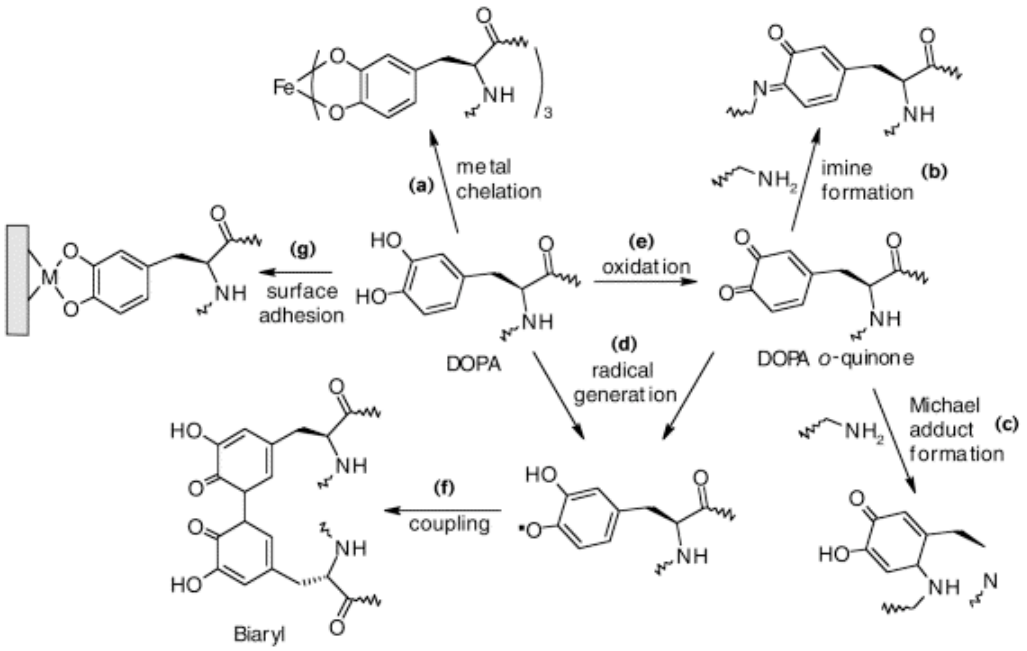


Figure 1. Hypothetical adhesion, oxidation, and cross-linking pathways for peptidyl DOPA residues and (e) DOPA *ortho*-quinone residues generation through oxidation. (a) Chelation of catechol by metal. (b) Coupling of DOPA-quinone with a pendant amine group to form an imine cross-link. (c) Coupling of DOPA-quinone with a pendant amino group in a Michael addition reaction to form a cross-link. (d) Reaction of DOPA or DOPA-quinone to form free-radical-containing species. (f) Coupling of DOPA-based radicals to form dimeric cross-linked products (biaryl). (g) Adsorption of DOPA to metal- (M) or metal-oxide bearing surfaces through hydrogen bonding or other means.

authors demonstrated solution-phase cross-linking of the polymers in the presence of the tyrosinase enzyme. DOPA-containing synthetic polypeptides have also been generated via the copolymerization of *N*-carboxyanhydride monomers of lysine and DOPA[34,35]. These copolymers were capable of forming hydrogels in the presence of oxidizing agents, including hydrogen peroxide and sodium periodate. In later work by the Schultz group, a site-specific strategy for incorporation of DOPA during *in vivo* protein synthesis was reported[37]. Genetic incorporation of the analogue was achieved using an orthogonal amber suppressor-tRNA/synthetase pair and employed for investigation of protein-protein interactions[37,38]. Both the *in vitro* and *in vivo* strategies described for unnatural amino acid incorporation, however, suffer from low protein yields that limit their use for synthesis of proteins with materials applications.

Work in our lab, however, has demonstrated the incorporation of L-DOPA into a mussel adhesive protein using a residue-specific strategy by replacing tyrosine with the unnatural amino acid analogue[39]. Herein, we describe work to extend the previous residue-specific incorporation strategy for the generation of an elastin-mimetic polypeptide containing periodically incorporated DOPA residues with a high degree of substitution. The *in vivo* incorporation approach should be characterized by improved yields. In addition, the newly synthesized elastin-mimetic polymer should be capable of undergoing efficient and versatile cross-linking via oxidation of the dihydroxy-phenylalanine side chain. Finally, the DOPA-containing polymer is an attractive substrate for biomedical materials applications such as the development of biocompatible surgical adhesives and sealants.

As a result of its chemical versatility, the unnatural amino acid DOPA has been useful for generation of polymer hydrogels through cross-linking of polymers comprising organic

monomers including those of polyethylene-glycol (PEG)[40-42] and polystyrene[43] to form mechanically robust hydrogels. While these studies have utilized oxidizers such as periodate to induce cross-linking, a more recent strategy has been inspired by the proposed mechanism of biosynthesis of native mussel byssus thread[44]. Recent evidence has supported that the presence of iron plays an important role in the adhesive properties of the proteins present within the byssus thread architecture[45]. Specifically, the assembly of the cuticle protein and its mechanical stability is in part due to the bonding of iron with the catechol-like amino acid L-DOPA to generate metal-ligand cross-links[45-47]. Based on the proposed metal-ligand cross-linking, Holten-Andersen and coworkers developed a strategy for introducing bis- and/or tris-catechol-Fe^{3+} cross-links into a synthetic polymer network derived from PEG polymers[44]. The generation and stoichiometry of the catechol-Fe^{3+} complexes is modulated by pH via oxidation of the hydroxyl groups of the unnatural amino acid residue (Figure 2). Using this strategy, the authors were able to generate a polymeric hydrogel that displayed high elastic moduli and self-healing properties.

The novel cross-linking strategy is acceptable for the design of biomaterials since the bis- and tris-catechol-Fe^{3+} complexes are characterized by higher stability than other metal-ligand chelates[48-51] and metal-DOPA bonds rupture under forces only modestly lower than covalent bonds[52]. In contrast to covalent bonds, the metal-DOPA bonds can spontaneously reform after rupture[52], which may provide a self-healing property to the native mussel adhesive protein[45]. We anticipate that biomaterials derived from elastin-mimetic polypeptides containing the unnatural amino acid DOPA will retain the advantages of other elastin-mimetic networks including biocompatibility and biostability. While the presence of the dihydroxy-phenylalanine side chain will permit formation of metal-ligand cross-links under physiologically relevant conditions. In

contrast to covalent cross-links, metal-DOPA cross-links within the elastin-mimetic polymer have the potential to reform after breaking and may impart a self-healing property to the generated hydrogel network. Finally, the residue-specific incorporation strategy should permit high level biosynthetic incorporation of the unnatural amino acid with high yields of the elastin-mimetic polypeptide.

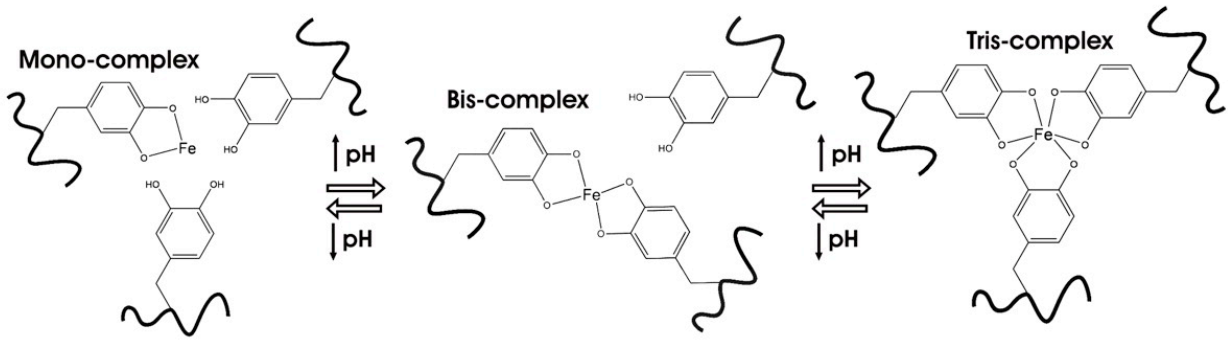
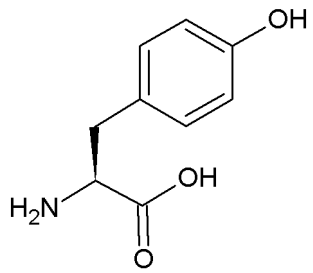
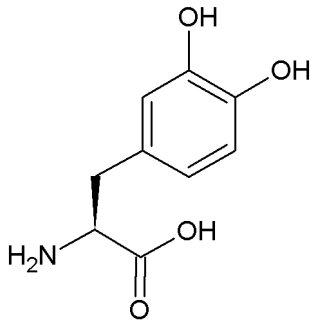


Figure 2. The Dopa-Fe^{3+} cross-linking mechanism inspired by the propose mechanism of cuticle protein cross-linking in mussels. The pH-dependent stoichiometry of the Fe^{3+}-catechol complexes.

(a)



L-tyrosine (**Tyr**)



L-3,4-dihydroxyphenylalanine (**DOPA**)

(b)

```
                  Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
                      ↓
AG CTT GAA GAC GTT CCA GGT GTT GGC GTA CCG GGT GTA GGC GTA CCA GGC
     A CTT CTG CAA GGT CCA CAA CCG CAT GGC CCA CAT CCG CAT GGT CCG
Hind III   Bbs I          ↑

Tyr Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
                                        ↓
TAC GGT GTA CCG GGT GTA GGC GTA CCG GGT GTT GGT GTT CCA AGA GAC GG
ATG CCA CAT GGC CCA CAT CCG TAT GGC CCA CAA CCA CAA GGT TCT CTG CCC TAG
                                                    ↑ BsmB I    BamH I
```

Scheme 1. (a) Chemical structures of the canonical amino acid (**Tyr**) and the tyrosine analogue (L-**DOPA**) that were used as substrates for biosynthetic incorporation into an elastin-mimetic polypeptide derived from concatemerization a (b) DNA monomer where the tyrosine codon represents the site of substitution with the analogue. (b) Oligonucleotide cassette encoding the **elastin-Y** repeat sequence. The recognition sites for the relevant type IIs restriction endonucleases, which are employed for generation of the DNA monomer, are highlighted. Arrows indicate the cleavage positions on the sense and anti-sense strand for the *Bbs* I and *Bsm*B I restriction endonucleases.

Experimental Methods

Materials

All chemical reagents were purchased from Fisher Scientific, Inc. (Pittsburgh, PA), Sigma-Aldrich Corporation (St. Louis, MO) or VWR International LLC (Radnor, PA) unless otherwise noted. Tyrosine analogue, L-3,4-dihydroxyphenylalanine (**L-DOPA**), was purchased from Research Organics, Inc. (Cleveland, OH). Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases were purchased from both New England Biolabs, Inc. (Beverly, MA) and Promega Corporation (Madison, WI), while antartic phosphatase, T4 DNA ligase, T4 polynucleotide kinase, and deoxynucleotide solution mix (dNTPs) were purchased only from New England Biolabs, Inc. Platinum® *Pfx* DNA polymerase, plasmid pZErO®-1 and DNA DipStick™ were obtained from Invitrogen Corp. (Carlsbad, CA).

Plasmid pQE-80L was purchased from QIAGEN Inc.-USA (Valencia, CA) while I-Lin Wu, a graduate student in the Conticello lab at Emory University, synthesized plasmids pIL2 and pIL5. Plasmid pPROTetE.133 was obtained from BD Biosciences, Inc. (Palo Alto, CA) and pPROLarA.231 was obtained from Professor Rik Myers of the University of Miami. Plasmid pHEC1 was previously synthesized by Dr. Holly Carpenter in the Conticello lab at Emory University[53]. Plasmid DNA preparation and purification were performed using the Mini-Prep 24™ (MacConnell Research Corporation, San Diego, CA), the QIAfilter Plasmid Maxi Kit, QIAprep Spin Miniprep Kit, QIAquick PCR Purification Kit, and QIAwell 8 Ultra Plasmid Kit (QIAGEN Inc., Valencia, CA), and the DNA Clean & Concentrator™ and Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA). Synthetic oligonucleotides were

purchased from either Sigma-Genosys (a division of Sigma-Aldrich Corporation, St. Louis, MO) or Integrated DNA Technologies (Coralville, IA) and were used as received. The *E. coli* strain, TOP10F′ was obtained from Invitrogen Corp. (Carlsbad, CA) and JW2581-1[54] from the *E. coli* Genetic Stock Center at Yale University (New Haven, CT; CGSC#: 10049). New minimal medium (NMM) was prepared according to the protocol of Budisa et al.[55] with the exception that L-tyrosine was not added to the medium prior to cell culture. The thermolysin protease from *Bacillus thermoproteolyticus rokko* was purchased as a lyophilized powder (~40 units/mg) from Sigma-Aldrich Corporation (St. Louis, MO).

General Methods

The molecular biology techniques utilized here, including expression cloning, polymerase chain reaction, gel electrophoresis, and growth and induction of bacterial cultures were adapted from a standard molecular cloning manual[56] or the protocol supplied by manufacturer, unless otherwise noted. Reagents for the manipulation of DNA, bacteria, and recombinant proteins were sterilized by either autoclave or passage through a syringe filter (0.2 μm cellulose membrane) or vacuum filter unit (standard polyethersulfone (PES) membrane) available from VWR International, LLC (Radnor, PA). Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. The concentration of DNA solutions was measured by an Ultrospec 3000 UV/Visible Spectrophotometer (Pharmacia Biotech, Cambridge, UK). Site-directed mutagenesis was performed using Stratagene's (La Jolla, CA) Quick-Change mutagenesis technique from gene-specific oligonucleotide primers. Polymerase chain reaction (PCR) was carried out using a GeneAmp 2400 Thermal Cycler (PerkinElmer Inc. Waltham, MA) and MJ Mini™ Gradient Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules, CA). Over the

course of these studies, automated DNA sequencing was performed at each of the following facilities: Beckman Coulter Genomics (Danvers, MA) on a Perkin-Elmer ABI Prism 377 DNA sequencer and GENEWIZ, Inc. (South Plainfield, NJ) on an Applied Biosystems 3730xl DNA analyzer. Agarose gel electrophoresis images were captured using a Kodak DC-120 digital camera or Kodak Gel Logic 112 imaging system from Carestream Health, Inc. (Rochester, NY). Protein electrophoresis was performed using 10-15% gradient discontinuous SDS PAGE pre-cast gels with a PhastSystem and visualized by silver staining using PlusOne Silver Stain Kit for proteins from GE Healthcare Bio-Sciences Corp. (Piscataway, NJ). Amino acid compositional analyses were performed at the W. M. Keck Foundation Biotechnology Resource Laboratory of Yale University (New Haven, CT) using the Hitachi L-8900 PH amino acid analyzer.

Construction of the elastin-Y gene

The **elastin-Y** multimer, which encodes a highly repetitive, elastin-mimetic polymer, was constructed using a previously reported concatemerization strategy[57]. The procedure for DNA cassette concatemerization was carried out as described in detail in Chapter 2 of this dissertation. The DNA oligonucleotide primers encoding the sense and anti-sense strands of the monomer sequence (Scheme 1) were chemically synthesized and annealed to produce duplex monomer DNA. The double stranded DNA cassette was identified via agarose gel electrophoresis (4% NuSieve®GTG® agarose, 1X TBE buffer) and purified by ethanol precipitation with Pellet Paint® co-precipitant. Double stranded DNA was phosphorylated through incubation with T4 polynucleotide kinase. The enzyme was removed by extraction with phenol/chloroform/isoamyl alcohol (25:24:1) and the double stranded DNA was recovered by ethanol precipitation.

The pZErO®-1 cloning plasmid (1 μg) was digested with *Bam*H I and *Hin*d III. The DNA

monomer insert and digested pZErO®-1 vector were ligated together in a corresponding 3:1 molar ratio in the presence of T4 DNA ligase (1 μL) at 16 °C for 30 min. An aliquot (2 μL) of the ligation mixture was used to transform electrocompetent cells of *E. coli* strain TOP10F' (40 μL). A total of 150 μL of the transformation recovery mixture was spread onto low salt LB agar media supplemented with 1 mM IPTG and Zeocin™ (50 μg/μL). Plasmid DNA from isolated from twelve single colonies using the QIAprep Spin Miniprep Kit. The recombinant plasmids were initially screened by double digestion with *Bam*H I and *Hin*d III. The DNA digestion products were visualized via agarose gel electrophoresis (4% NuSieve®GTG® agarose, 1X TBE buffer) for the presence of the DNA monomer insert. The sequence of the elastin-Y monomer gene was confirmed by automated DNA sequencing analysis using the M13 reverse and M13 (-20) forward primers (Appendix 1). Plasmids containing the correct insert sequence were designated **pMAP28** and stored at –20 °C.

A large amount of the DNA monomer was required for DNA cassette concatemerization. The **pMAP28** plasmid DNA was isolated from four-2.8 L flasks containing 500 mL low salt LB supplemented with Zeocin™ (50 μg/mL) using the QIAfilter Plasmid Maxi Kit. The monomer DNA (~500 μg) was liberated from the cloning plasmid by sequential digestion with *Bbs* I and *Bsm*B I and purified via agarose gel electrophoresis (2% NuSieve®GTG® agarose, 1X TBE buffer) using the Amicon Ultrafree MC Maximum Recovery Kit (Millipore; Burlington, MA) followed by ethanol precipitation.

Self-ligation of the DNA monomer in a head-to-tail fashion, or multimerization, was performed by incubating the monomer DNA in the presence of T4 DNA ligase (10 μL) and T4 DNA ligase buffer at 16 °C for 16 h. The library of DNA concatemers was fractionated via

agarose gel electrophoresis (1% agarose, 0.5X TBE buffer). The region corresponding to the desired multimer size range (~2500 to 1500 bp) was excised from the gel and purified using the Zymoclean™ Gel DNA Recovery Kit and quantified using DNA Dipstick™. The pool of DNA concatemers was then ligated into a *Bbs* I-digested and dephosphorylated **pMAP28** acceptor plasmid. An aliquot of the ligation mixture was used to transform electrocompetent cells of *E. coli* strain TOP10F'. A total of 150 μL of the transformation recovery mixture was spread onto low salt LB media agar supplemented with 1 mM IPTG and Zeocin™ (50 μg/μL). The plates were incubated for 12 to 14 h at 37 °C. Twelve single colonies were selected and plasmid DNA was isolated using the automated Mini-Prep 24™. The recombinant plasmids were initially screened via double digestion with *Bam*H I/*Hin*d III and analyzed via agarose gel electrophoresis (1% agarose, 0.5X TBE buffer) for the presence of multimer inserts.

From the pool of transformants, a colony was isolated containing a recombinant plasmid with a concatemer insert in the desired size range of 1500 to 2500 bp. The plasmid was used to transform fresh electrocompetent TOP10F' cells and isolated from the bacteria cells using the QIAprep Spin Miniprep Kit. The sequence of the recombinant plasmid generated by concatemerization of the elastin-Y gene was confirmed via automated DNA sequencing analysis using the M13 reverse and M13 (-20) forward primers. The correctly sequenced plasmid was designated **pMAP32** (Figures 3 and 4) and stored at –20 °C.

Cloning of elastin-Y gene into expression plasmid

The genes encoding the elastin-mimetic polypeptides were cloned into the commercially available pQE-80L expression vector. Use of the seamless cloning strategy required insertion of an adaptor sequence containing recognition sites for type IIs restriction endonucleases. An

expression plasmid was prepared via insertion of a modified polylinker sequence into the commercially available pQE-80L to generate the plasmid **pIL2**. For subsequent cloning of the elastin-Y concatemer, it was necessary to remove a *Bsa* I restriction endonucleases recognition site from the modified expression plasmid. Site-directed mutagenesis was utilized to introduce a silent mutation into the sequence of the plasmid and produce the plasmid **pIL5** (*vide ante*). The modified cloning plasmid **pIL5** contained a copy of the overproducing repressor allele *lacI*^{q} to ensure tight control of the basal level of transcription prior to induction with IPTG.

The elastin-mimetic DNA concatemer was then cloned into the modified pQE-80L expression vector **pIL5**. The **pMAP32** plasmid was sequentially digested with *Bbs* I and *Bsm*B I to isolate the multimer gene. The **pIL5** plasmid was digested with *Bsa* I to generate the complementary sticky ends necessary for seamless cloning of the concatemer gene into the adaptor sequence. Prior to ligation, the linearized plasmid DNA was also dephosphorylated with antartic phosphatase to reduce the incidence of false positive ligations. The concatemer elastin-Y insert and pQE-80L vector DNA fragments (~100 ng of each) were incubated with T4 DNA ligase (1 μl) and T4 DNA ligase buffer at 16 °C for 12 h. An aliquot (2 μL) of the ligation reaction was used to transform electrocompetent cells of TOP10F' (40 μL). The recovery mixture (150 μL) was plated onto LB agar media supplemented with ampicillin (100 μg/mL). The plates were incubated at 37 °C for 12 to 14 h. Six single colonies were selected from each plate for screening. The recombinant plasmid DNA was isolated from the cultured cells under ampicillin selection using the QIAprep Spin Miniprep Kit. The DNA was screened via double digestion with *Eco*R I and *Hin*d III for the presence of the elastin-Y concatemer insert. The sequence of the expression plasmid was confirmed by automated DNA sequencing analysis

using primers designed to bind the pQE-80L plasmid upstream and downstream of the adaptor sequence (Appendix 1). The correctly sequenced plasmid **pMAP35** (Figure 4) includes the modified pQE-80L plasmid with insertion of the ~1800 bp DNA concatemer encoding the **elastin-Y** polymer.

Synthetase plasmid construction

Construction of a plasmid encoding the wild-type *E. coli* tyrosyl-tRNA synthetase (TyrRS) was previously described[39]. Briefly, the hybrid vector **pME1**[58] was employed as the expression vector for the TyrRS cassette. A recombinant copy of the aminoacyl-tRNA synthetase gene, *tyrS* (Appendix 2), was amplified from the *E. coli* genomic DNA using gene specific primers, tyrS-F (5'-AGGTTGGTACCATGGCAAGCAGTAACTT GATTAAACAATTGC-3') and tyrS-R (5'-GGAGCTCTAGATTATTTCCAGCAAATCAGACAGTAATTCTTTTTACCG-3'). The amplification primers were designed to introduce *Kpn* I and *Xba* I restriction endonuclease cleavage sites at the 5' and 3' termini of the amplified gene, respectively. Genomic DNA was isolated from *E. coli* strain TOP10F' cells using the Dneasy Tissue Kit and protocol (QIAGEN, Inc.). The acceptor plasmid **pME1** and amplified *tyrS* gene were double digested with *Kpn* I and *Xba* I and the digestion products were purified via agarose gel electrophoresis (1% agarose, 0.5X TBE buffer). The digested TyrRS DNA cassette was ligated with the digested, dephosphorylated **pME1** acceptor plasmid using T4 DNA ligase via incubation at 14 °C for 14 h. Chemically competent cells of *E. coli* TOP10F' strain were transformed with an aliquot (2 μL) of the ligation mixture. Transformants were plated on LB agar media supplemented with chloramphenicol (35 μg/mL) and tetracycline (10 μg/mL) for 16 h at 37 °C. Single colonies were used to inoculate LB (5 mL) media supplemented with chloramphenicol

and incubated overnight in a rotating drug at 37 °C. The plasmid DNA was isolated using the QIAprep Spin Miniprep Kit and screened via double digestion with *Kpn* I and *Xba* I for the presence of an insert of size concomitant with the TyrRS cassette. Automated DNA sequencing analysis with primers designed to bind the pPROTet/Lar plasmid upstream downstream of the multiple cloning site was used confirm the identity of the recombinant clone (Appendix 1). The plasmid encoding an additional copy of the wild-type *E. coli* tyrosyl-tRNA synthetase plasmid was designated **pJA7** (Figure 6).

Bacterial growth and expression

Small scale expression in JW2581-1 strain

Chemically competent cells of the tyrosine auxotrophic *E. coli* strain JW2581-1[54] [F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, Δ*tyrA763::kan, rph-1, Δ*(rhaD-rhaB)568, hsdR514*] were prepared. Cells were cotransformed with the plasmid **pMAP35**, which encodes the **elastin-Y** polymer, and the **pJA7** plasmid (containing the recombinant wild-type *E. coli* TyrRS cassette) or with the plasmid **pHEC1**[53] (a modified pPROTet/Lar plasmid lacking the synthetase gene). Aliquots (150 μL) of the recovery mixture were plated on LB agar media supplemented with ampicillin (100 μg/mL), chloramphenicol (35 μg/mL), and kanamycin (30 μg/mL). Glycerol stocks of the expression strains were prepared from 800 μL of overnight liquid cell culture and 200 μL of 80% glycerol and stored at –80 °C.

Five milliliters of LB media supplemented with ampicillin, chloramphenicol, and kanamycin, as required for plasmid and strain maintenance, were inoculated with single colonies of the JW2581-1 expression strains and grown overnight at 37 °C on a rotating drum. One milliliter (2% of the final cell culture volume) of the overnight culture was used to inoculate new

minimal media (NMM; 7.5 mM $(NH_4)_2SO_4$, 8.5 mM NaCl, 55 mM KH_2PO_4, 100 mM K_2HPO_4, 1 mM $MgSO_4$, 20 mM glucose, 1 mM $CaCl_2$, 10 mg/mL thiamine hydrochloride, 10 mg/mL biotin in PBS, and 50 μg/mL of each of the 20 canonical amino acid expect tyrosine) supplemented with antibiotics and 0.4 mM L-tyrosine (prepared as a 200 mM stock solution in 1X phosphate buffered saline, PBS at pH 10) for a final culture volume of 50 mL. The fifty-milliliter cultures were incubated at 37 °C with agitation at 225 rpm until the optical density (OD) measure at 600 nm reached 0.8 to 1.0 (approximately 2 to 3 h). The cells were harvested by centrifugation at 4 °C, 4000 x *g* for 10 min. The supernatant was discarded and the cell pellets were gently suspended in 10 mL aqueous solution of sterile ice-cold (0 °C) sodium chloride (0.9% NaCl). The cell suspensions were centrifuged at 4 °C, 4000 x *g* for 10 min. The supernatant was discarded and the cell pellet was washed a second time in the isotonic sodium chloride solution. The supernatant was again discarded and the cell pellets were suspended in 60 mL NMM without addition of antibiotics or tyrosine. The cell cultures were then divided into 6-10 mL aliquots and incubated at 37 °C with shaking at 225 rpm to deplete the remaining cellular levels of tyrosine.

Tyrosine and the unnatural amino acid analogue (2*S*)-2-amino-3-(3,4-dihydroxyphenyl) propanoic acid (L-**DOPA**) were prepared fresh as 200 mM stock solutions in 1X PBS with titration of 1 M NaOH until soluble and stored at 4 °C. Aliquots of the tyrosine or unnatural amino acid stock solution were added to the 10 mL cell cultures for a final concentration of 1 mM tyrosine and 0.5, 1.0, 2.0, or 4.0 mM L-**DOPA**. An aliquot (10 μL) of a 1 M IPTG stock solution (prepared by dissolving 596 mg IPTG in 2.5 mL sterile ddH_2O followed by filter sterilization) was added to the cultures for a final concentration of 1 mM to induce expression of

the elastin-mimetic polypeptides. The cultures were incubated at 37 °C with shaking at 175 rpm for 3 h after induction. One-milliliter aliquots of the cell cultures were collected at 3 h and centrifuged at 4 °C for 5 min at 6300 rpm. The supernatant was discarded and the cell pellets were suspended in 30 μL sterile ddH_2O and stored at –20 °C for further analysis by protein electrophoresis. The whole cell lysates were analyzed by SDS PAGE and the accumulation of protein was visualized by silver stain. Briefly, whole cell lysates were prepared by mixing aliquots (5 μL) of the suspended cell pellet samples with 12.5 μL 2X SDS gel-loading buffer (100 mM Tris-Cl, pH 6.8; 4 % (w/v) SDS, electrophoresis grade; 0.2% (v/v) bromophenol blue; 20% (v/v) glycerol; 2.5 μL 1 M DTT, and sterile ddH_2O to a final volume of 25 μL. The mixtures were then boiled at 100 °C for 5-10 min and cooled on ice an additional 10 min, prior to gel loading.

Large scale (1 L) expression

Plasmids **pMAP35** (modified pQE-80L expression plasmid encoding the **elastin-Y** polymer) and **pJA7** (containing a recombinant copy of the wild-type *E. coli* tyrosyl-tRNA synthetase gene) were cotransformed into the JW2581-1 strain and plated on LB agar media supplemented with ampicillin (100 μg/mL), chloramphenicol (35 μg/mL), and kanamycin (30 μg/mL). Twenty milliliters of LB media supplemented with antibiotics were inoculated with single colonies of the JW2581-1 transformants. The cells were grown overnight at 37 °C with shaking at 225 rpm. The overnight seed cultures were harvested at 4 °C, 4000 x *g* for 20 min. The supernatant was discarded and the cell pellets were suspended in 20 mL NMM.

The cell suspensions were divided into 2-10 mL aliquots (2% of total culture volume) and each was used to inoculate a 2-2.8 L flasks (each containing 490 mL NMM supplemented with

ampicillin (100 μg/mL), chloramphenicol (35 μg/mL), and kanamycin (30 μg/mL) for a total culture volume of 500 mL). In this way, two flasks were used for elastin-mimetic biosynthesis from 1000 mL (1 L) for both expression strain types. The flasks were incubated at 37 °C with shaking at 225 rpm to an $OD_{600\ nm}$ of 0.8 to 1.0 over a period of approximately 2 to 3 h. Once the cultures reached log phase growth, the intact cells were harvested by centrifugation of the 500 mL cultures at 4 °C, 4000 x *g* for 10 min. The supernatant was discarded and the cell pellets were washed with an aqueous solution of sterile ice-cold (0 °C) 0.9% NaCl (2 x 100 mL). The cell pellets were finally suspended in 500 mL of NMM lacking tyrosine but supplemented with antibiotics, ampicillin (50 μg/mL), chloramphenicol (35 μg/mL), and kanamycin (10 μg/mL), at reduced concentrations to maximize protein expression. The cell cultures were incubated at 37 °C for 30 min with shaking at 225 rpm to deplete the remaining cellular levels of tyrosine.

The tyrosine analogue L-3,4- dihydroxy-phenylalanine (L-**DOPA**) was prepared fresh as a 200 mM stock solution in PBS with titration of 1 M NaOH for a final pH of ~12 and used immediately. An aliquot of 2.5 mL of tyrosine or the L-**DOPA** stock solutions (200 mM) was added to each of the 500 mL cell cultures for a final concentration of 1 mM. An aliquot (500 μL) of a 1 M IPTG stock solution was added to the cultures for a final concentration of 1 mM to induce elastin-mimetic protein expression. The cultures were incubated at 37 °C with shaking at 175 rpm for 3 h after induction. The cells were then harvested by centrifugation at 4 °C, 4000 x *g* for 20 min. Cell pellets from the 2-500 mL cultures of the same expression strain were combined and suspended in 50 mL lysis buffer (50 mM sodium phosphate, 300 mM NaCl, pH 5.5) and stored at –80 °C.

Purification of the elastin-Y polymers

Purification of the **elastin-Y** and **elastin-DOPA** containing polymers was carried out under mildly acidic conditions (pH 5.5) to reduce oxidation of the unnatural amino acid side chain to a quinone, which could potentially induce cross-linking or sclerotization reactions. The pH was selected to also preserve activity of the enzyme employed for cell lysis. Three freeze (-80 °C)/ thaw (25 °C) cycles were employed for initial cell fracture. Lysozyme from chicken egg white was added to the cell lysate at a final concentration of 2 mg/mL (from a 10 mg/mL stock solution) along with 1X Protease Inhibitor Cocktail Set I (Calbiochem®, a brand of EMD Chemicals, Gibbstown, NJ). The mixture was incubated at room temperature for 30 min with agitation. Benzonase® nuclease and 1 M $MgCl_2$ were then added to a final concentration of 25 units/mL and 1 mM, respectively. The solutions were incubated at 4 °C for 36 h.

The cold cell lysate was centrifuged at 20,000 x *g* for 40 min at 4 °C to pellet the cellular debris. 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 while some residual protein remained in the pellet. The elastin-mimetic polymers were then purified via inverse transition cycling according to a protocol adapted from Meyer and Chilkoti[59]. The elastin-mimetic polypeptides were salted out of solution by addition of 5 M NaCl to a final concentration of 2 M NaCl and incubation at 30 °C for 60 min without agitation. Aggregation of the elastin-mimetic polymers was observed as evidenced by an increase in turbidity of the solution. The solutions were centrifuged at 30 °C, 10,000 x *g* for 15 min to recover the aggregated proteins (‘hot-spin’). The supernatant was discarded and the protein pellet was suspended in ice-cold (0 °C) lysis buffer (~ 42 mL) with

addition of 1X Protease Inhibitor Cocktail Set I. The proteins were suspended on ice for only 30-45 min to avoid solubilizing unwanted contaminates. A second cycle of centrifugation was performed ('cold-spin') at 4°C, 20,000 x *g* for 15 min. The supernatant was transferred to a sterile tube and salting out was repeated with 2 M NaCl at 30 °C for 60 min. The hot- (30 °C)/ cold- (4 °C) spin cycles were repeated until no contaminating pellet was observed following a cold spin. Typically, three cycles were sufficient to purify the elastin-mimetic polypeptides to homogeneity. Following the final hot-spin, the resulting protein pellets were suspended in ice-cold (0 °C) lysis buffer (~10-15 mL total volume) without addition of the protease inhibitor cocktail. The suspended protein pellets were incubated at 4 °C with shaking overnight. The elastin-mimetic protein solutions were dialyzed at 4 °C using the SnakeSkin® Pleated Dialysis Tubing with a molecular weight cut-off of 10 kDa (Pierce Protein Research Products, Thermo Fisher Scientific, Inc.; Rockford, IL) against distilled, deionized water (5 x 4 L). The protein solutions were further purified via sterilization at 4 °C through a 5.0 μM Durapore® membrane filter (Millipore; Billerica, MA) using a Fisherbrand® glass microanalysis filter holder assembly. Lyophilization of the filtered dialysate produced the elastin-mimetic polypeptides as white spongy solids. The lyophilized proteins were analyzed by 10-15 % gradient discontinuous SDS PAGE (PhastSystem) and protein bands were visualized by silver staining (PlusOne Silver Stain Kit).

Physical and analytical measurements

Mass spectrometry of thermolysin digested polypeptides

Dilute aqueous solutions of the elastin-mimetic polypeptides were prepared as 1 mg/mL in buffer (50 mM sodium phosphate, 300 mM NaCl, pH 5.5) at room temperature (25 °C) to

retain polymer solubility. The thermolysin protease solution was prepared as an aqueous (1 mg/mL) solution with 2 mM Ca^{2+} (2 mM $CaCl_2$). The enzyme was added to the polymer solutions for a final 1:50 ratio of enzyme: protein, and the mixtures were incubated at room temperature (25 °C) overnight. The protease digestion reaction was purified using the Pierce® C18 Spin Columns (Pierce Protein Research Products, Thermo Fisher Scientific, Inc., Rockford, IL) and the polypeptide fragments were eluted with 70% acetonitrile: 0.1% TFA. The organic solvent was removed by brief evaporative centrifugation using the Jouan Inc. RC10-10 centrifugal evaporator (Winchester, VA). The collected peptide fragments were suspended in 20 μL of 0.1% formic acid.

Thermolysin products were analyzed by LC-MS. Digestion products were separated by liquid chromatography using an Ascentis C18 HPLC column (3 μm particle size, 3cm x 2.1 mm) from Sigma-Aldrich, Corp (St. Louis, MO). The bound peptides were eluted with 5-50 % (v/v) acetonitrile: 0.1% TFA at a flow rate of 0.2 mL/min. The protease cleavage fragments of interest eluted early in the acetonitrile gradient (approximately 5%) and thus, associated weakly with the column. Electrospray ionization of the peptides was carried out in positive ion-mode using an LTQ FT Ultra Hybrid Mass Spectrometer (Thermo Fisher Scientific, Inc., Waltham, MA). The mass spectral data were processed into peak lists with Analyst QS (Applied Biosystems) and compared to the theoretical *m/z* of the predicted protease cleavage sequences.

Nuclear magnetic resonance (NMR) spectroscopy

Solution NMR spectra of the **elastin-Y** polymers were acquired using a Varian UNITY Plus 600 instrument (600 Mhz ^1H) from Agilent Technologies (Santa Clara, CA). The ^1H-NMR spectra were collected at 4 °C on samples consisting of 10 to 15 mg of protein dissolved in 0.7

mL deuterium oxide (D, 99%) from Cambridge Isotope Laboratories, Inc. (Andover, MA). Chemical shifts for the ^{1}H nuclei were referenced and reported relative to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS)[60]. The NMR data were processed using the ACD/NMR Processor Academic Edition software from ACD/Labs (Ontario, Canada).

Temperature-dependent turbidity

The thermal transitions of the **elastin-Y** polymers were estimated from temperature turbidimetry profiles. Measurements were performed using a JASCO V-530 UV/Visible spectrophotometer equipped with a programmable water-controlled Peltier cell and a JASCO ETC-505T temperature controller (Jasco, Inc., Easton, MD). Dilute solutions of the elastin-mimetic polymer samples (1.0 mg/mL) were prepared at 4 °C and analyzed in a quartz cuvette with a 10 mm path length (Hellma Analytics; Müllheim, Germany). The absorbance of the copolymer solutions was monitored over a temperature range of 4 °C to 70 °C at a wavelength of 256 nm with a ramp rate of 1 °C/min and rescan of the samples was performed *in situ* following a 15 min pre-scan thermostat. Spectra were recorded and plotted using the Temperature-Scan (Melting) analysis feature of the JASCO Spectra Manager software package.

Metal-ligand binding of elastin-DOPA polymer

The oxidation of the DOPA side chain and formation of the DOPA-Fe^{3+} catechol complexes was monitored using the JASCO V-530 UV/Visible spectrophotometer equipped with a programmable water-controlled Peltier cell and a JASCO ETC-505T temperature controller (Jasco, Inc.; Easton, MD). Spectral changes of the elastin-mimetic polymer solutions were measured at 4 °C using a quartz cuvette with a 10 mm path length (Hellma Analytics; Müllheim, Germany). The elastin-mimetic polymer solution (1 mg/mL **elastin-DOPA$_{21}$**, 0.44 mM DOPA)

was prepared at 4 °C and 1.8 μL of an 80 mM $FeCl_3$ aqueous solution was added to give a 3:1 molar ratio of DOPA: Fe^{3+}. The absorbance of the solution was monitored at increasing pH with titration of 1 M NaOH to a final pH of ~12. The absorbance from 350 nm to 950 nm was plotted using the JASCO Spectra Manager software package.

Formation of an elastin-DOPA hydrogel

A five hundred-microliter Fe^{3+}-catechol cross-linked hydrogel was prepared. Twenty-five milligram of **elastin-DOPA** polymer was dissolved in 250 μL sterile ddH_2O at 4 °C (DOPA concentration of 22 mM). The polymer solution was mixed with 1/6 final volume (83.3 μL) of 80 mM (13 mg/mL) $FeCl_3$ on a sterile watch glass. Formation of the gel was completed by adding 2/6 final volume (166.7 μL) of 1 M NaOH. The final pH (~12) was selected to drive formation of the tris-catechol-Fe^{3+} complexes within the elastin-mimetic polymer. The gel was mixed at 4 °C an additional 30-60 s until a homogeneous color and physical state were established.

Results and Discussion

The elastin-mimetic DNA cassette **elastin-Y** (Scheme 1) was designed to encode an artificial protein where, under selective pressure, the natural amino acid tyrosine is replaced by the unnatural analogue L-**DOPA** during *in vivo* protein synthesis. A previously reported seamless cloning strategy[57] was employed to generate a library of DNA concatemers encoding the elastin-Y polymer. The polymer comprises a highly repetitive sequence derived from the consensus pentapeptide repeat motif of native elastin (Val-Pro-Gly-Xaa-Gly). Cloning of the DNA concatemer library was performed using the cloning plasmid pZErO®-1, which contains the smaller Zeocin™ resistance gene reducing the overall vector size for increased transformation and ligation efficiencies. From DNA cassette concatemerization, a recombinant clone containing a multimer of approximately 1800 bp was isolated (Figure 3). The recombinant plasmid designated **pMAP32** (Figures 4) was utilized for subsequent cloning experiments. The DNA concatemer was removed from the cloning plasmid and inserted into a modified pQE-80L expression vector to generate plasmid **pMAP35** (Figure 4). The plasmids utilized and constructed in the synthesis of the **elastin-Y** polymers are described in Table 1. The IPTG-inducible expression system was utilized to direct synthesis of an elastin-mimetic polypeptide; where the non-canonical amino acid DOPA was globally incorporated at the fourth position of the pentapeptide repeat periodically along the polypeptide chain. In this way, the amino acid composition, length, sequence, and density of incorporation sites are controlled at the molecular level by the amino acid sequence.

Based on previous work in our lab[39], we hypothesized that the structurally similar analogue of tyrosine, L-**DOPA**, will serve as a substrate for the host strain translational

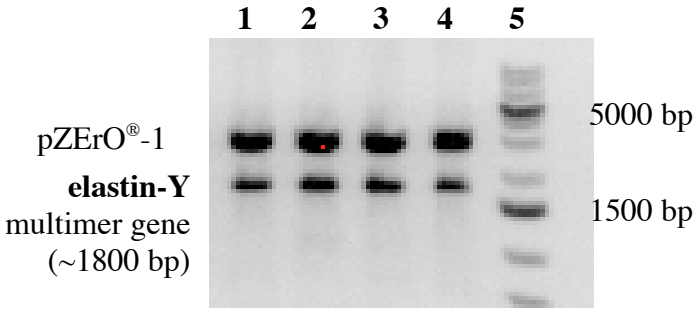


Figure 3. A 1% agarose gel depicting plasmid **pMAP32** (~500 ng) double digested with *Bam*H I and *Hind* III indicating the **elastin-Y** multimer gene insert. Twenty microliters of the digestion reaction were run on the gel in separate lanes. Lane 1-4: The DNA from 4 single colonies transformed with the correctly sequenced plasmid **pMAP32**, which contains the pZErO®-1 plasmid and **elastin-Y** multimer insert; and lane 5: O'GeneRuler™ 1 kb Plus DNA ladder (Fermentas Inc.; Glen Burnie, MD).

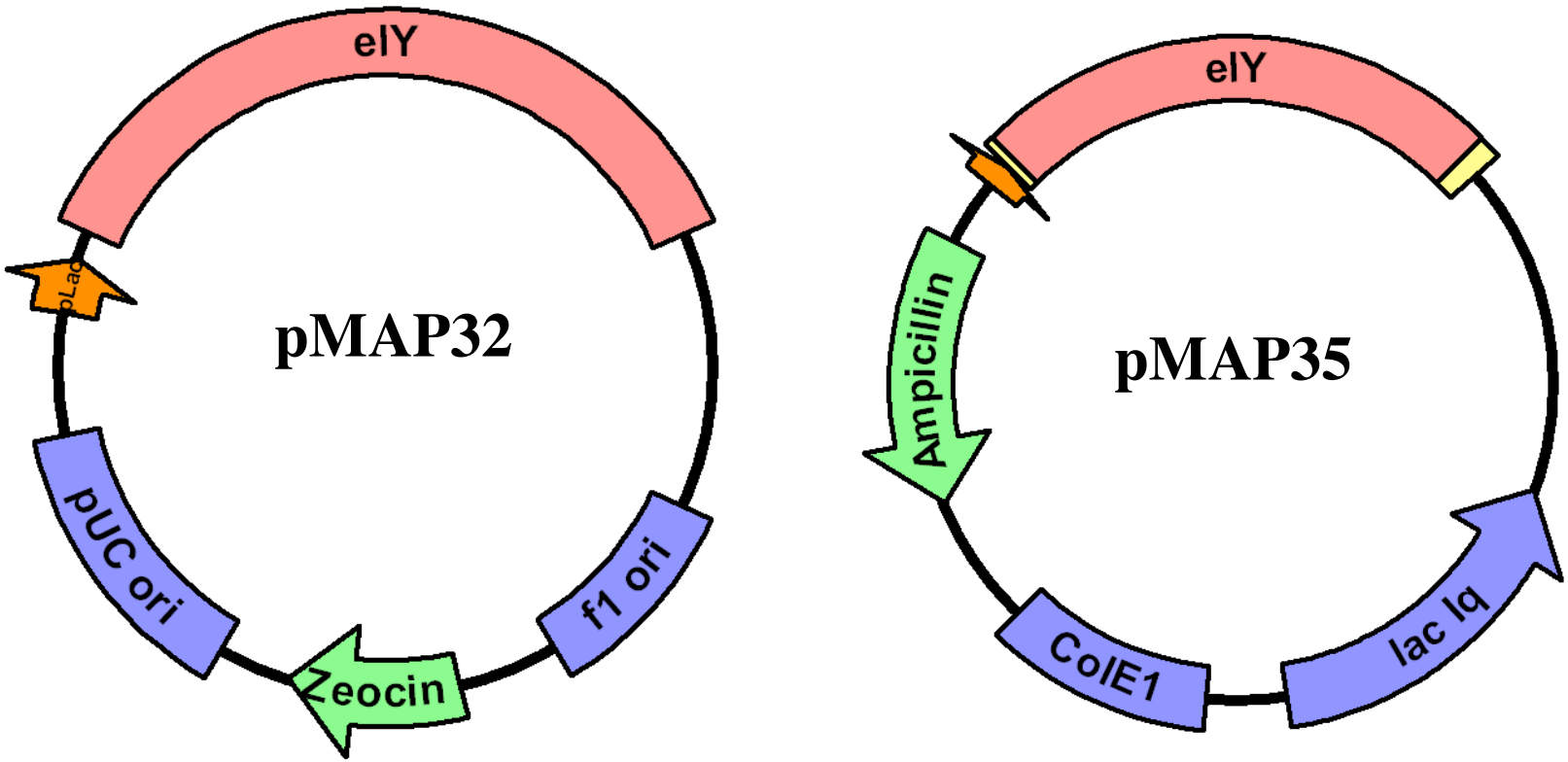


Figure 4. Plasmid **pMAP32** represents the recombinant pZErO®-1 plasmid following recloning of the pool of elastin-Y concatemers in the desired size range of 1500 to 2500 bp. Plasmid **pMAP35** depicting the modified pQE-80L expression plasmid encoding the **elastin-Y** polymer.

Table 1. List of recombinant plasmids utilized and constructed in the synthesis of the **elastin-Y** polymers.

Recombinant Plasmid Name	Plasmid/Insert	Comments
pMAP28	pZErO®-1/**elastin-Y** DNA monomer	Kn^R
pMAP32	pZErO®-1/**elastin-Y** concatemer	
pIL2	pQE-80L/**pQE Adaptor**	Amp^R
pIL5	pIL5 with removal of internal *Bsa* I site	
pMAP35	pIL5/**elastin-Y**	
pME1	pPROTetE.133/LarA.231	Cm^R, p15A ori; Reference[58]
pJA7	pME1/*tyrS*	*tyrS* sequence, Appendix 2; Reference [39]
pHEC1	Modified pME1 lacking aaRS	Reference [53]

machinery. Using a residue-specific strategy, we investigated incorporation of the unnatural amino acid within the elastin-mimetic polypeptide chain in response to tyrosine codons. Incorporation of the unnatural amino acid at multiple sites within the **elastin-Y** polypeptide was investigated using the tyrosine auxotroph JW2581-1 (*ΔtyrA763::Kan*), characterized by knockout of the *tyrA* gene involved in tyrosine biosynthesis[54], under conditions of tyrosine depletion and analogue supplementation.

Previous work in our lab, investigated a residue-specific incorporation method for replacement of tyrosine by the unnatural amino acid DOPA in a mussel adhesive mimetic protein[39]. The results of these experiments demonstrated that coexpression of the wild-type tyrosyl-tRNA synthetase (TyrRS) was not necessary for protein expression in the presence of L-DOPA, although coexpression of the TyrRS was required for incorporation of the more structurally divergent analogue of tyrosine 3-amino-L-tyrosine. In accordance with these results, we examined synthesis of **elastin-Y** in the auxotrophic strain JW2581-1 under conditions of tyrosine depletion and analogue supplementation without coexpression of the TyrRS. Evidence of protein expression under these conditions supported incorporation of DOPA at a level of efficiency similar with synthesis in the presence of the canonical amino acid (positive control) (Figure 5). Consistent with previous studies of elastin-mimetic polypeptides, the molar mass of the elastin-mimetic polypeptides as observed by SDS PAGE appear to be ~20% larger than the calculated masses[16,61].

Previous work by Tirrell and coworkers has investigated the incorporation of methionine analogues with increased activity of the methionyl-tRNA synthetase[62]. The authors found that the kinetics of activation of the analogues by the wild-type aaRS correlates well with the level of

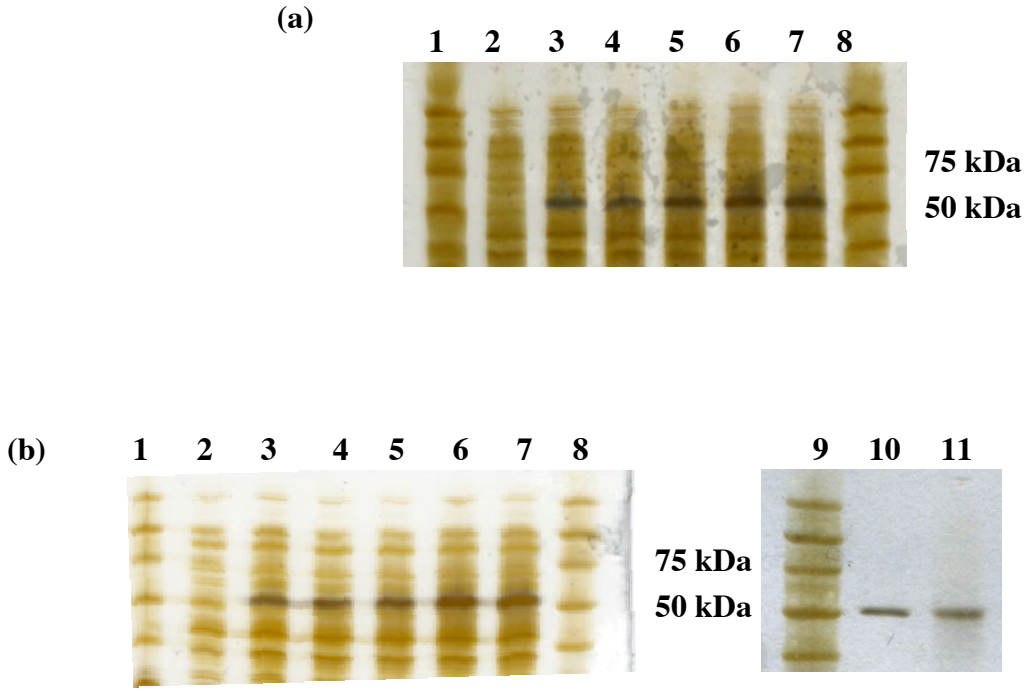


Figure 5. SDS PAGE with silver staining indicating multi-site incorporation of L-3,4-dihydroxyphenylalanine (L-**DOPA**) within the elastin-mimetic polymer **elastin-Y**. Elastin-mimetic polypeptides were synthesized from *E. coli* strain JW2581-1 cotransformed with the plasmid encoding the artificial protein **pMAP35** and a modified pPROTet/Lar hybrid plasmid (a) **pHEC1** (empty plasmid) or (b) **pJA7** (plasmid encoding an additional copy of the wild-type tyrosyl-tRNA synthetase). Whole cell lysates from expression cultures were analyzed after 3 h induction with 1 mM IPTG. Lane 1,8,9: Perfect Protein™ Marker, 15-150 kDa (Novagen®, a brand of EMD Chemicals, Inc., Gibbstown, NJ); lane 2: tyrosine deficient (negative) control; lane 3: tyrosine supplemented (1 mM; positive) control; lane 4-7: L-**DOPA** supplemented (0.5, 1.0, 2.0, and 4.0 mM, respectively); lane 10: purified proteins from cultures supplemented with tyrosine (1 mM); and lane 11: purified proteins from cultures supplemented with L-**DOPA** (1 mM). The expected molecular weight for the elastin-mimetic polypeptides is approximately 47 kDa.

protein synthesis supported by the analogues *in vivo*. The authors show that by manipulating the activity of the aaRS through use of a modified host strain overexpressing the synthetase, the yield of proteins containing the methionine analogues could be improved. It follows that overexpression of the aminoacyl-tRNA synthetase may provide a general strategy for improving and optimizing yields of proteins containing other non-natural amino acids. Thus, for DOPA incorporation experiments we also employed a plasmid **pJA7** (Figure 6), which contains a recombinant copy of the wild-type *E. coli* tyrosyl-tRNA synthetase gene. The TyrRS expression cassette is placed under control of the P_L*tet* promoter. In the *E. coli* auxotrophic strain JW2581-1, which is unable to synthesize the *tet* repressor, the synthetase is expressed constitutively from the respective plasmid **pJA7**. Electrophoretic analysis of whole cell lysates derived from small scale expression cultures indicated that the analogue displayed a level of target protein expression comparable to positive control, in which cells were provided the canonical amino acid tyrosine (Figure 5). These results confirm that coexpression of the wild-type aminoacyl-tRNA synthetase was not necessary for protein production from cultures supplemented with the analogue L-**DOPA**.

The **elastin-Y** derivatives were isolated from large scale (1 L) cultures using the inverse temperature cycling protocol previously described[59]. The isolated yields of the polypeptides (Table 2) ranged from approximately 16 mg/L to 32 mg/L for fully induced expression cultures in minima media (NMM) supplemented with 1 mM of the appropriate amino acid under conventional batch fermentation conditions in shake flask culture. The protein is isolated in relatively high yields, especially in consideration of the high density of sites for analogue incorporation in the target polypeptide sequence (Scheme 2). Protein yields of **elastin-DOPA**

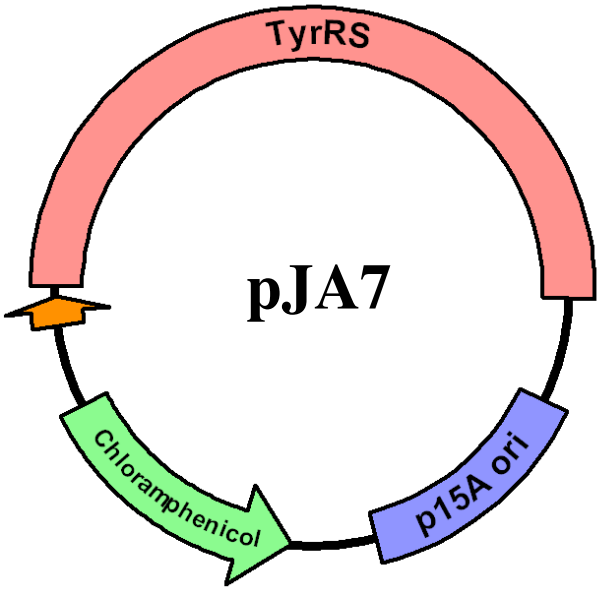


Figure 6. The plasmid **pJA7** was generated by cloning of the PCR amplified TyrRS gene into the *Eco*R I and *Hin*d III sites of a previously synthesized pPROTet/Lar hybrid plasmid, **pME1**[58].

Scheme 2. Amino Acid Sequence of Protein-Based Block Copolymer **elastin-Y**

elastin-Y M-[X]-VPGVGGSD$_4$KGH$_{10}$,

where [X] = [(VPGVG)$_2$VPGYG(VPGVG)$_2$]$_{21}$

Table 2. Expression yields for elastin-mimetic polypeptides from large scale expression (1 L). Host strain JW2581-1 was cotransformed with plasmid encoding target elastin-mimetic polypeptide and the **pJA7** (TyrRS expression) plasmid or the **pHEC1** (control) plasmid supplemented with 1 mM amino acid as indicated.

Elastin-mimetic polypeptide	Coexpression pJA7	Yield (mg/L)
elastin-Tyr	**No**	24.4.0
elastin-DOPA	**No**	16.6
elastin-Tyr	**Yes**	32.4
elastin-DOPA	**Yes**	21.4

from the host strain overexpressing TyrRS were moderately improved over yields from cells expressing an empty (control) plasmid. This result is consistent with the hypothesis that overexpression of aaRS can result in increase yield of proteins containing non-natural amino acids.

The identities of the elastin-mimetic polymers were investigated by mass spectrometry. Initial attempts at using MALDI-TOF MS, however, were unsuccessful at producing a spectrum for the full-length **elastin-DOPA** that estimated the molecular mass of the polymer with confidence. To assay the success of analogue incorporation, the highly repetitive proteins (Scheme 2) were digested into fragments of similar composition using the protease thermolysin. The masses of the resulting fragments were analyzed by LC-MS. The most abundant fragment predicted by digestion of **elastin-Tyr** with thermolysin comprises the peptide (VGVPGYGVPG) with a theoretical mass of 901.48 Da. The fragment is observed in the LC-MS spectrum following digestion of **elastin-Tyr** and the observed *m/z* agrees with the exact theoretical mass (Figure 7) predicted for a single charged species. The other peaks observed in the full ESI-MS spectrum may be due to buffer salts that were not removed during C18 column purification or fragments of the enzyme. Digestion of **elastin-DOPA** reveals an analogous fragment that has *m/z* in agreement with substitution by the unnatural amino acid (Figure 8). The spectrum also contains a much smaller peak with *m/z* ratio corresponding to the theoretical *m/z* for a peptide cleavage fragment containing the canonical amino acid tyrosine. These results suggest that tyrosine replacement by the unnatural amino acid occurs with a high degree of biosynthetic substitution but is not complete. Current efforts are underway to determine the molecular mass of the full-length protein.

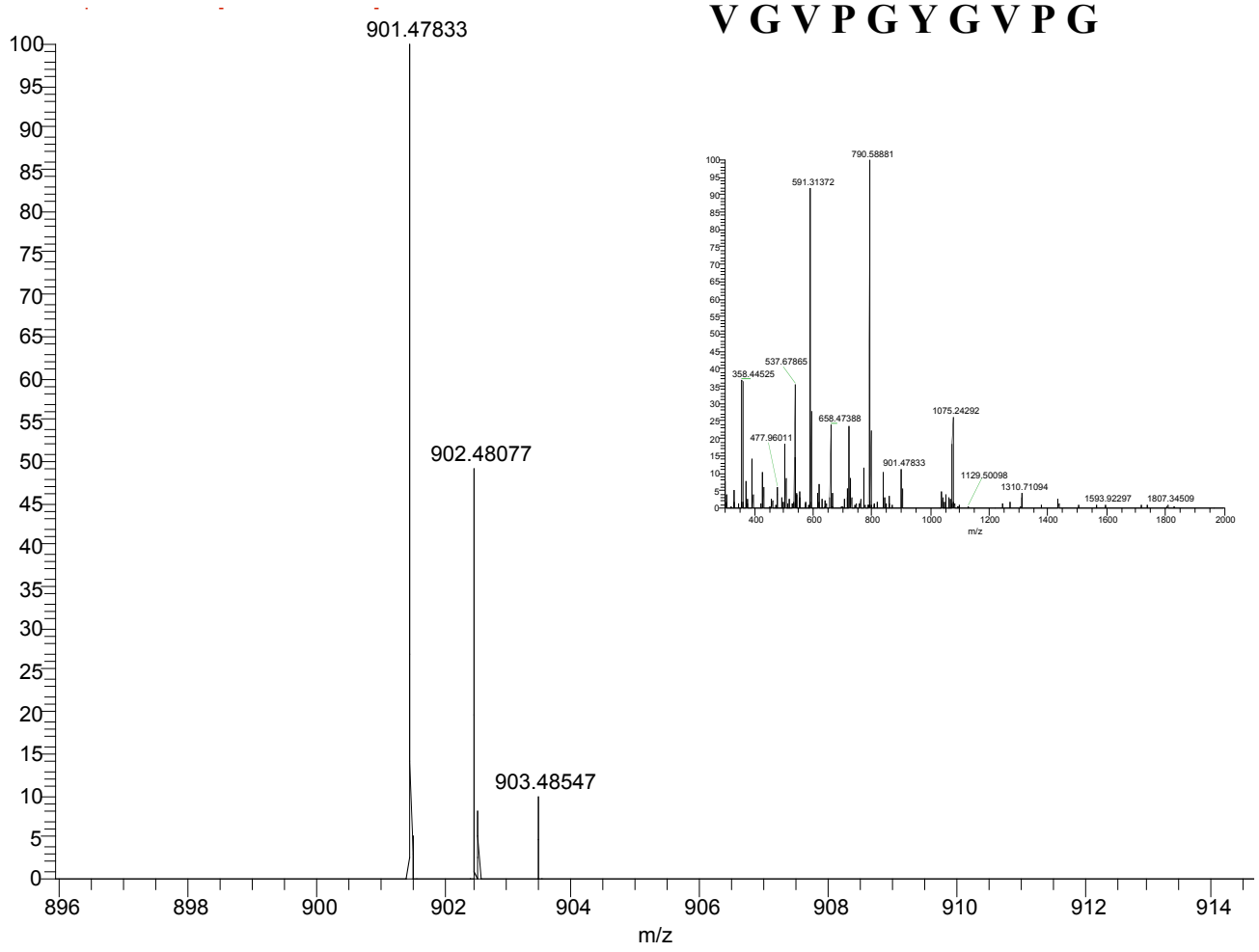


Figure 7. LC-MS reveals a fragment of **elastin-Tyr** generated by protease digestion with thermolysin. The observed *m/z* for this fragment (901.48) agrees exactly with the theoretical *m/z* (901.48) for the predicted digestion product. The full ESI-MS spectrum is also shown (inset).

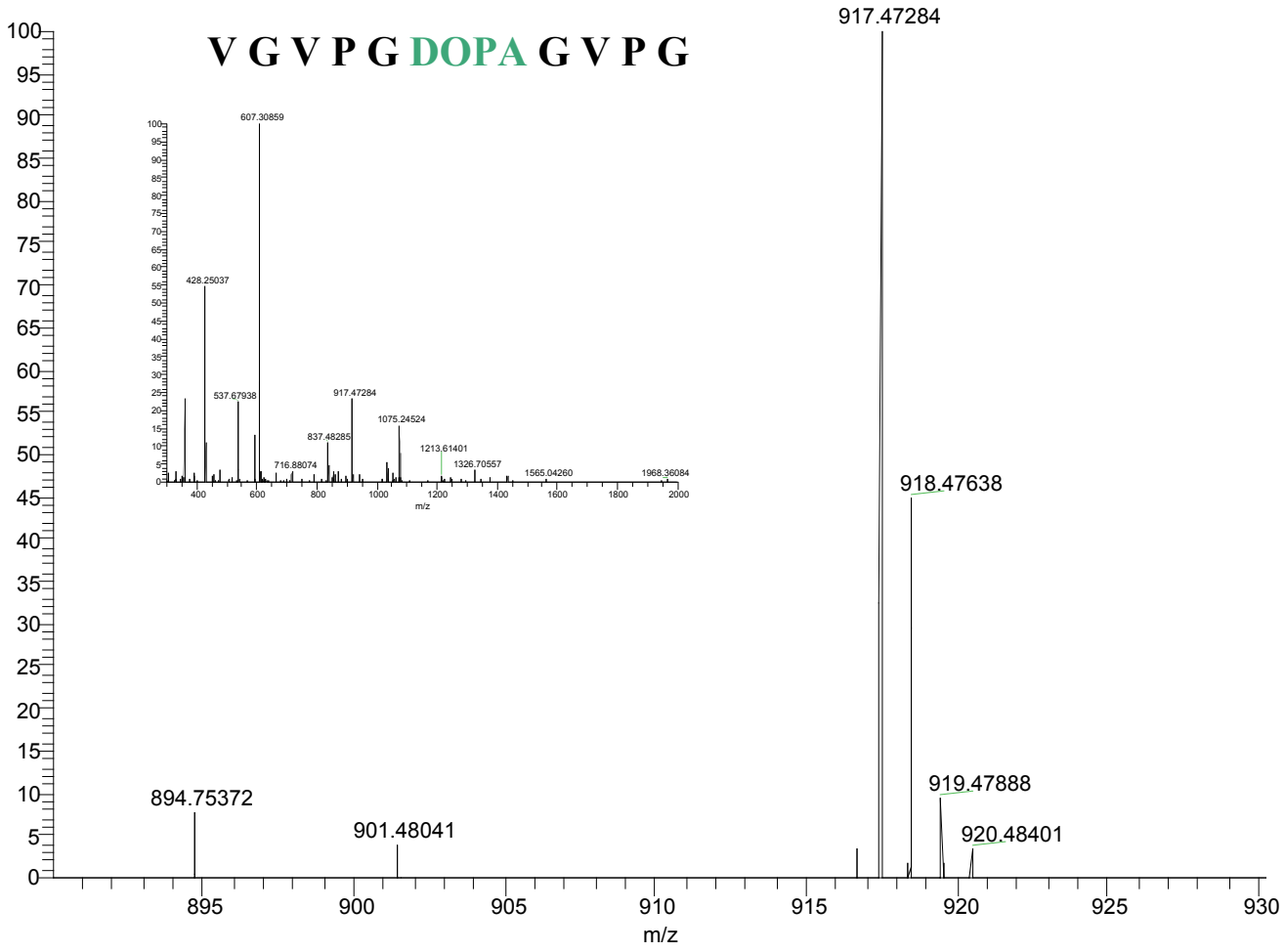


Figure 8. LC-MS reveals a fragment of **elastin-DOPA** generated by protease digestion with thermolysin. The observed *m/z* for this fragment (914.47) agrees exactly with the predicted *m/z* (917.47) for the fragment with replacement of tyrosine in the **elastin-Y** sequence by the unnatural amino acid DOPA (Scheme 1). The full ESI-MS spectrum is shown (inset).

Amino acid analysis of the purified **elastin-DOPA** indicated that the protein contained 5.2 mol% L-**DOPA** (Table 3). The observed mole percent of the unnatural amino acid is greater than the predicted value (3.8 mol%), however, direct quantification of the analogue is difficult possibly because the amino acid is partially destroyed during the HCl hydrolysis protocol. The analysis did estimate that the **elastin-DOPA** protein contained only 0.2 mol% of tyrosine, which indicates a significant decrease in the tyrosine content compared to the **elastin-Tyr** polymer (3.7 mol%). Assuming that the diminution in tyrosine content is due to replacement by DOPA, the percent substitution with the unnatural amino acid analogue could also be calculated as a percentage of the missing tyrosine. These calculations indicated incorporation of the unnatural amino acid within the **elastin-DOPA** polymer at 94.4%.

Further evidence for incorporation of the DOPA amino acid analogue into the polypeptide was obtained from ^{1}H-NMR spectroscopy (Figure 9). Spectroscopic analysis of the **elastin-DOPA** polymer demonstrated structural features that were commensurate with high levels of incorporation of the unnatural amino acid analogue. Comparison of the regions of the ^{1}H-NMR spectra corresponding to the aromatic side chain of the elastin-mimetic polymers clearly shows the presence of a singlet corresponding to the CH resonance at the C-2 position of the dihydroxy-phenylalanine ring in **elastin-DOPA**. In the spectrum of the **elastin-DOPA** polymer, a small doublet is observed at a chemical shift of 6.5 ppm, which suggests incomplete replacement of tyrosine. Integration of the doublet peak, however, indicates that analogue incorporation efficiency can be estimated at >90%. Taken together, the LC-MS, amino acid compositional analysis, and NMR spectroscopy data support that L-**DOPA** is an acceptable substrate during *in vivo* protein synthesis in an *E. coli* expression host.

Table 3. Amino acid compositional analysis of the purified **elastin-Tyr** and **elastin-DOPA** polymers.

Amino Acid	**elastin-Tyr**	**elastin-DOPA**
	Calculated (Observed) [mol-%]	
Asparagine (D)	0.7 (0.7)	0.7 (0.8)
L-DOPA		3.8 (5.2)
Glycine (G)	39.1 (39.8)	39.1 (41.3)
Histidine (H)	1.8 (0.6)	1.8 (1.0)
Lysine (K)	0.2 (0.1)	0.2 (0.2)
Methionine (M)	0.2 (0.2)	0.2 (0.2)
Proline (P)	19.3 (20.1)	19.3 (20.0)
Serine (S)	0.2 (0.1)	0.2 (0.1)
Tyrosine (Y)	3.8 (3.7)	0.0 (0.2)
Valine (V)	34.9 (34.0)	34.9 (34.5)

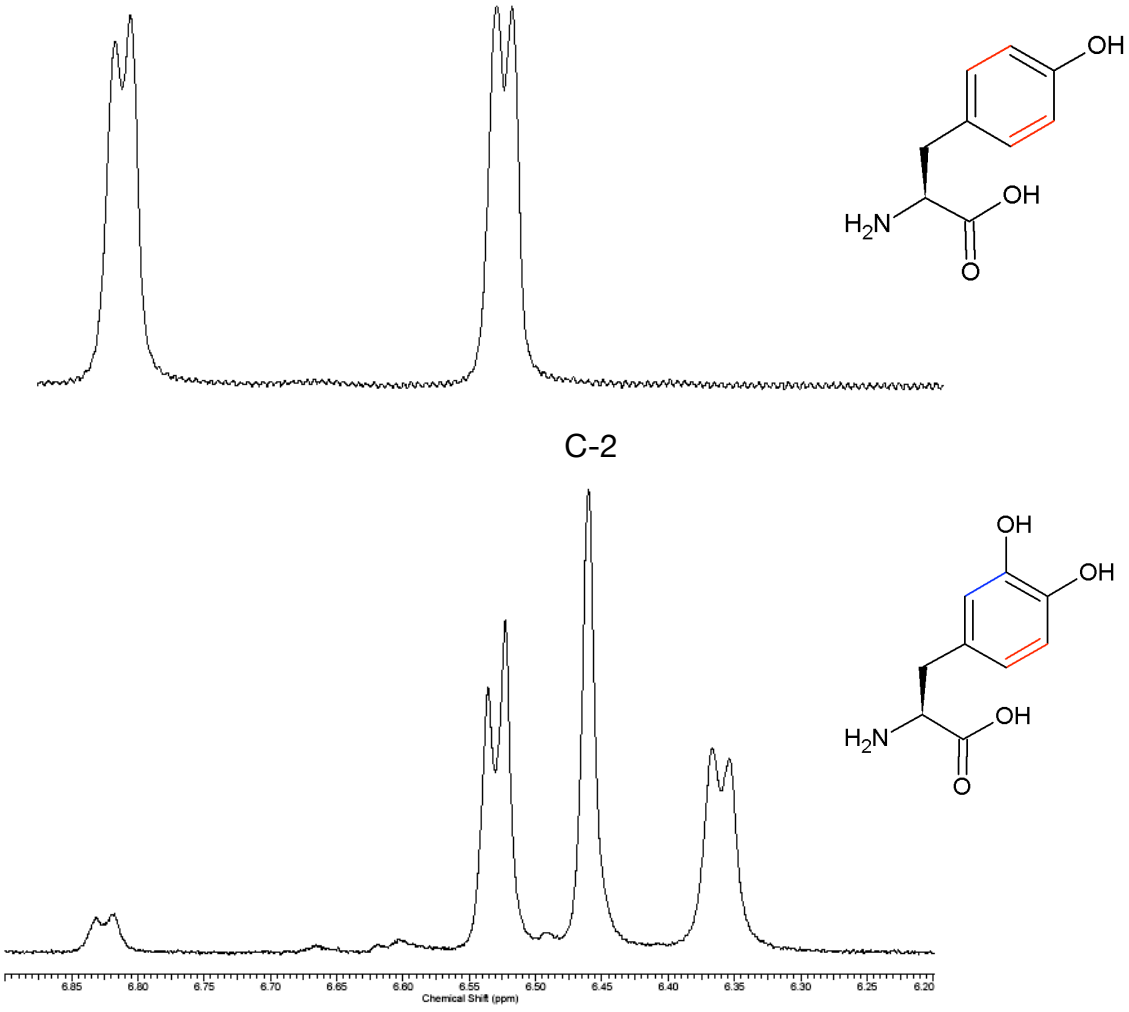


Figure 9. NMR spectroscopic analyses of the elastin derivatives. ^{1}H-NMR spectra of the **elastin-Tyr** (upper spectrum) and **elastin-DOPA** (lower spectrum) polymers along with the chemical structures of the corresponding amino acids provided during protein expression.

In our study, the expression conditions were found compatible for expression of an elastin-mimetic protein variant with a high degree of biosynthetic substitution.

Temperature turbidity experiments were conducted on the elastin-mimetic polymer substituted with tyrosine and DOPA and monitored using UV-visible spectrophotometry. The thermal transition of the polymers is observed by an increase in absorbance. Cooling and rescan of the samples *in situ* demonstrate the reversibility of the transitions. Under acidic conditions, the **elastin-Tyr** and **elastin-DOPA** polymers undergo temperature-responsive aggregation over a narrow temperature range of approximately 20 °C to 25 °C (Figure 10). The temperature maximum of the inverse phase transitions is consistent with the presence of the nonpolar residues in the fourth position of the pentapeptide. At neutral pH, the **elastin-Tyr** polymer solution undergoes a reversible coacervation characterized by a sharp transition from the soluble to the aggregate phase over a temperature with a midpoint of approximately 20 °C (Figure 11). In contrast, dilute aqueous solutions of **elastin-DOPA** at neutral pH undergo aggregate formation over a broad temperature range of approximately 15 °C to 25 °C. Furthermore, aggregation of the polypeptides is not completely reversible as evidenced by an absorbance value above the baseline at low temperatures (10 °C) following rescan of the samples. Even after prolonged thermostat at 4 °C, the solutions remained slightly turbid (data not shown). The temperature turbidimetry profiles for **elastin-DOPA** indicate that at neutral pH the elastin-mimetic polypeptides may form strong aggregates above the T_t, which are unable to disassemble upon cooling. Oxidation of the DOPA residue side chain to a quinone may induce cross-linking that is not observed under acidic conditions. Finally, neither the **elastin-Tyr** nor **elastin-DOPA** polymers

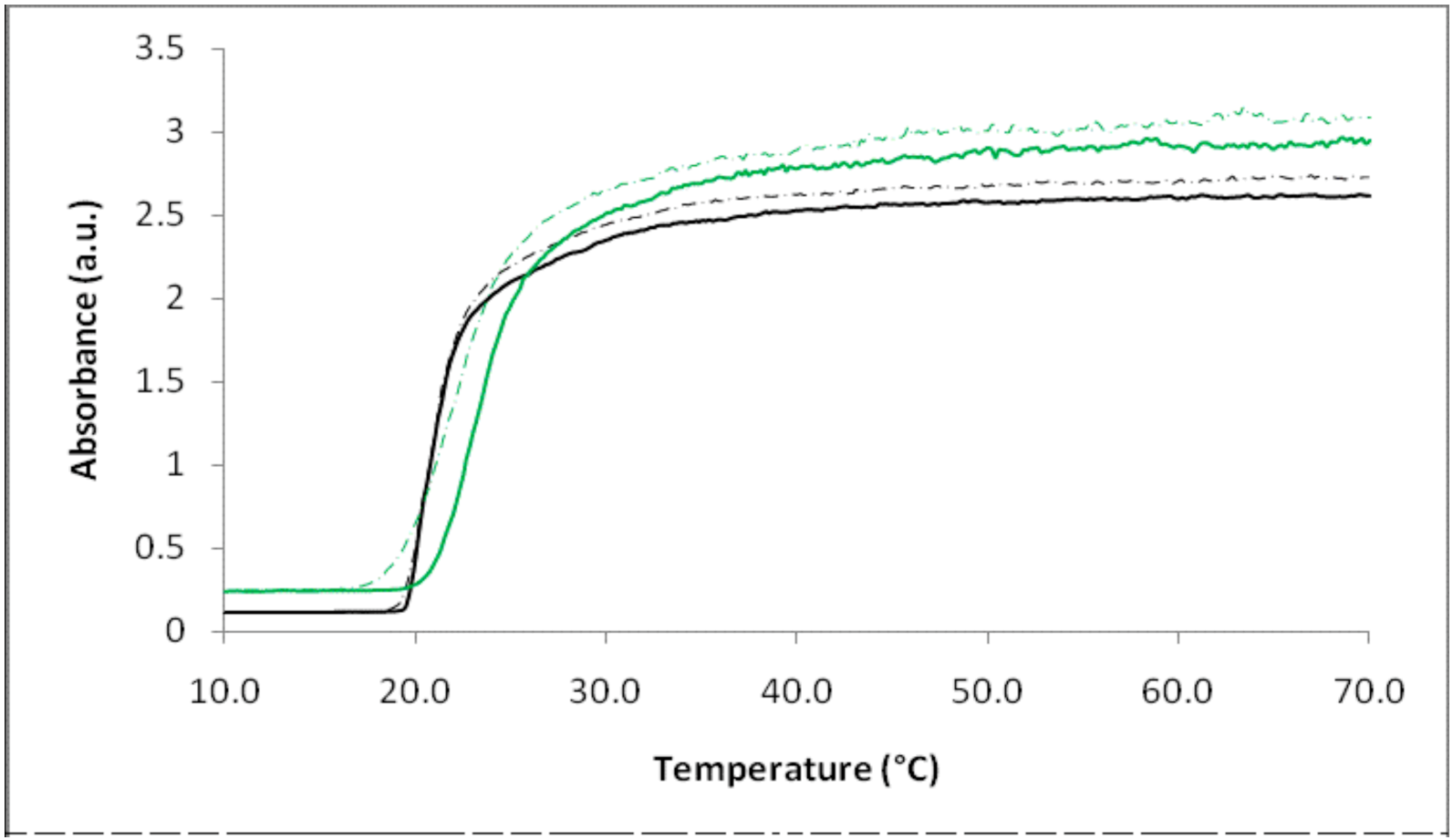


Figure 10. Temperature-dependent turbidimetry profiles for solutions of the **elastin-Tyr** (shown in black) and **elastin-DOPA** (shown in green) in dilute aqueous solutions at low pH (40 mM acetic acid). The reversibility of the transitions is demonstrated by rescan of the polymer solutions (dotted lines).

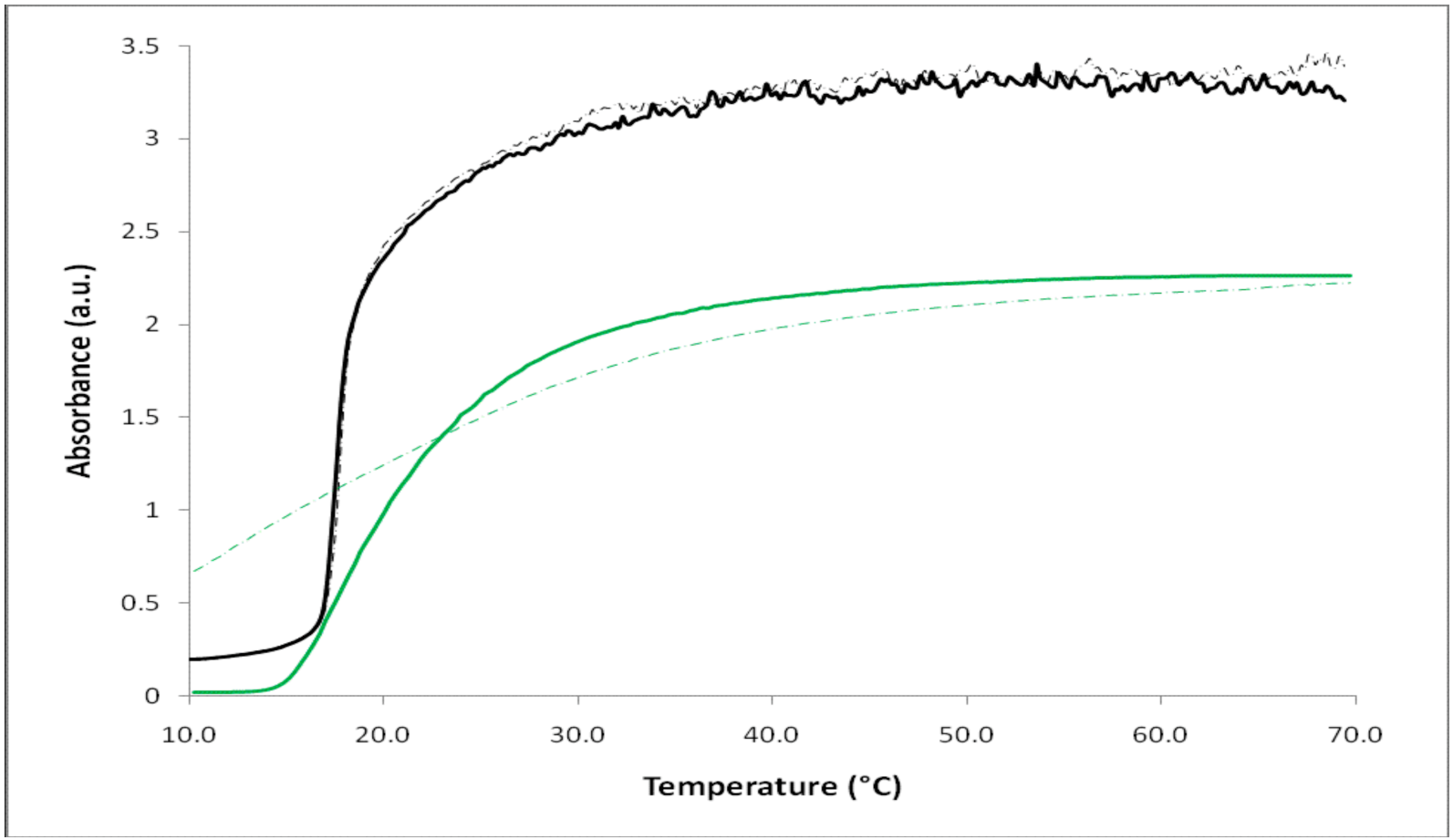


Figure 11. Temperature-dependent turbidimetry profiles for solutions of the **elastin-Tyr** (shown in black) and **elastin-DOPA** (shown in green) in dilute aqueous solutions at neutral pH. The reversibility of the transitions is demonstrated by rescan of the polymer solutions (dotted lines).

form a coacervate under basic conditions (40 mM NaOH) at temperatures below 70 °C, which indicates that deprotonation of the hydroxyl groups inhibits the hydrophobic self-assembly process. The observed absence of the inverse phase transition below 70 °C upon deprotonation of the **tyrosine** and L-**DOPA** residues coincides with the predicted effect of increased backbone polarity on the coacervation temperature of elastin-mimetic polymers. Specifically, Urry and coworkers have described the chemical synthesis of a related series of elastin-mimetic polypeptides derived from the consensus pentapeptide sequence (Val-Pro-Gly-Xaa-Gly), and monitored the effect of substitution at Xaa on the observed temperature of the inverse phase transition[63]. For polypeptides encoding tyrosine at position Xaa, the T_t was shifted to high temperatures above the pKa of the tyrosine side chain (pH = 10), which resulted in deprotonation of the hydroxyl moiety. In summary, the newly generated elastin-mimetic polymer **elastin-DOPA** incorporating the unnatural amino acid DOPA periodically within the protein backbone forms a coacervate, which is held together via hydrophobic forces. The temperature of the inverse phase transition is near ambient room temperature, and the newly synthesized elastin-mimetic polypeptide may serve as a useful module for the generation of an amphiphilic block copolymer capable of both physical and chemical cross-linking.

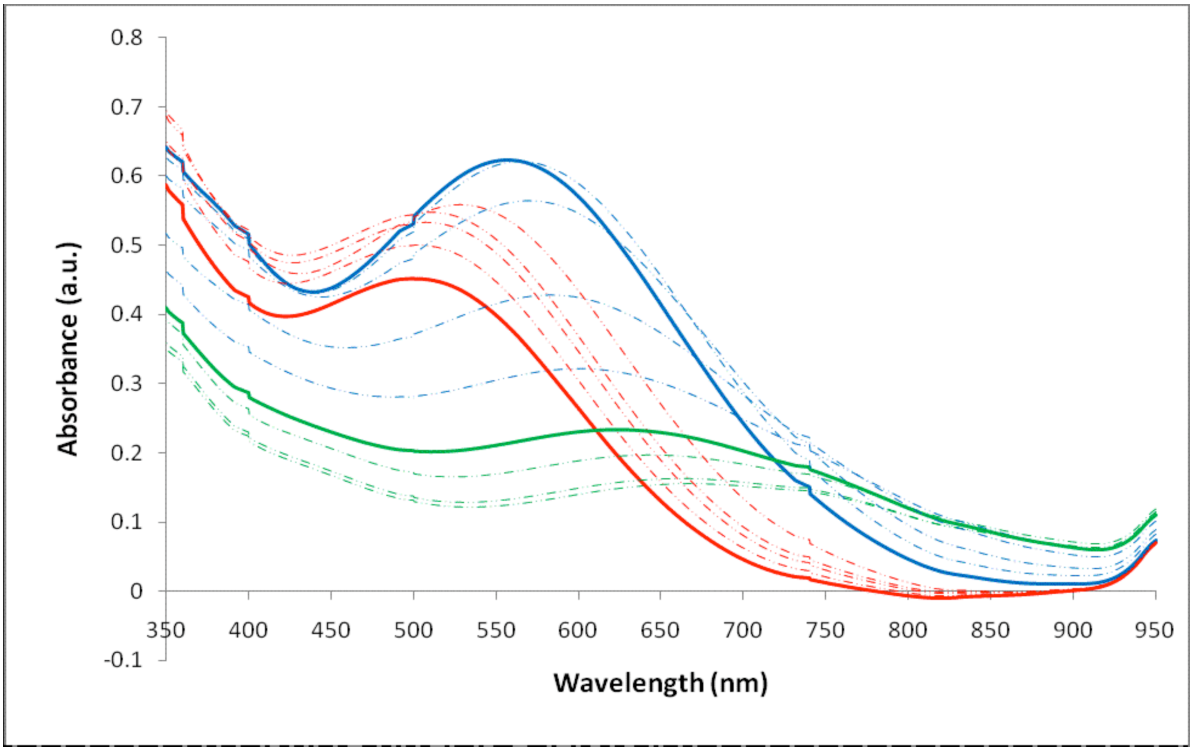
Fabrication of chemically cross-linked elastomeric matrices have relied upon a wide variety of methods including γ radiation-induced[64], photoinitiated[7,65], and chemical cross-linking[5,6,8-18,66] as well as enzymatic-based approaches[67]. For use in biomaterials applications, many of these approaches are met with limitations such as the lack of chemoselectivity, the cytotoxicity of the photo-cross-linking catalyst or chemical cross-linker, and the occurrence of unintended side reactions. Incorporation of DOPA presents a new mechanism for chemically

cross-linking elastin-mimetic polypeptides into elastomeric hydrogel matrices. A recently reported strategy[44] was investigated which introduces cross-links between DOPA residues via metal-ligand binding. The cross-linking mechanism was inspired by the process of native byssus thread formation by mussels, where DOPA plays a role in the adhesive properties of the proteinaceous cuticle through chelation of iron to form metal-ligand complexes[45,46]. Holten-Andersen and coworkers investigated the formation of catechol-Fe^{3+} polymer cross-links by titration of a solution of PEG polymers, containing a four-residue DOPA endgroup, and $FeCl_3$ with 1 M NaOH to induce oxidation of the dihydroxy-phenylalanine side chain. The authors found that the PEG-$DOPA_4$ polymers were able to form tris-catechol-Fe^{3+} cross-links resulting in a synthetic polymer network and that the resulting material displayed high elastic moduli and self-healing properties. Based on the success of the pH-induced metal-ligand cross-linking strategy for polymers derived from conventional organic monomers, we applied the method to the newly synthesized **elastin-DOPA**. The proposed cross-linking mechanism is advantageous over previous methods of generating covalently cross-linked networks of elastin-mimetic polypeptide. Cross-linking can be performed under biologically relevant conditions, the nature of the cross-linked is well defined and controllable via pH, and the cross-link density can be modulated through selection of the amino acid sequence, chain length, and sites of unnatural amino acid incorporation.

To investigate the potential for **elastin-DOPA** to form a polymer network through the metal-ligand cross-linking strategy, we first monitored the stoichiometric transitions of the catechol-Fe^{3+} complexes. Maintaining a temperature of 4 °C, dilute solutions (~1 mg/mL) of **elastin-DOPA** were mixed with $FeCl_3$ in a 3:1 molar ratio of DOPA: Fe^{3+} and the pH of the

polymer solution was increased by titration with 1 M NaOH. The spectra changes from formation of the mono-, bis-, and tris-catechol-Fe^{3+} complexes within **elastin-DOPA** were monitored by UV-visible spectrophotometry (Figure 12a). The initially green solution turned blue and finally wine red with increasing pH as the catechol-Fe^{3+} stoichiometry was altered from mono-, bis-, and tris-, respectively. The DOPA: Fe^{3+} stoichiometry is estimated from previous studies of model peptides[68]. The UV-visible spectra was interpreted and correlated well with the previous spectroscopy studies, indicating predominance of the dimer by a peak at 575 nm and the trimer by a peak at 492 nm. Further, the characteristic peaks in the spectra corresponding to the dominant species (mono-, bis-, or tris-catechol:Fe^{3+}) were observed in pH ranges in agreement with previous reports[44]. Finally, we tested the effect of the catacheol-Fe^{3+} cross-linking on concentrated polymer networks. The elastin-mimetic polymer was prepared as a 5-wt% solution in sterile ddH_2O with a stoichiometric excess of $FeCl_3$ and the solution was mixed gently at 4 °C. A large volume of 1 M NaOH was added to induced complete oxidation of the DOPA residue side chains and drive the equilibrium towards formation of the tris-catechol-Fe^{3+} complexes. Upon increasing the pH to approximately 12, the wine red gel was immediately formed and an additional 30-60 s gentle agitation resulted in a gel of homogeneous color (Figure 12b). Additional characterization of the hydrogel architecture and resulting materials properties should be conducted on the newly synthesized **elastin-DOPA** to investigate its potential for biomaterials application.

(a)



(b)

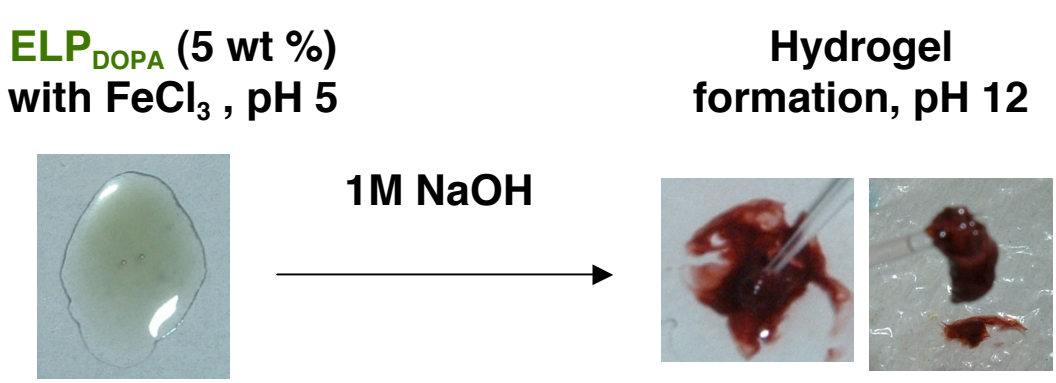


Figure 12. (a) UV-visible absorbance of **elastin-DOPA**$_{21}$-Fe^{3+} solution monitored with increasing pH prepared from 1 mg/mL **elastin-DOPA** with $FeCl_3$ (DOPA: Fe molar ratio of 3:1) with titration of 1 M Na OH. All green curves are at pH < 5, all blue curves are in the range of 5 < pH < 9.5 where the bis-complex dominates, and all red curves are at pH > 10 where the tris-complex dominates. (b) Formation an **elastin-DOPA** hydrogel from a 5-wt% solution. In the presence of $FeCl_3$, the green solution forms a wine red gel upon increasing the pH. The gel was lifted off the watch glass (right-side photo).

Conclusion

In summary, we have established an *E. coli* expression strain that is competent for incorporation of the unnatural amino acid DOPA and developed culture conditions that result in a high degree of substitution of the analogue within an artificial protein. The method has been employed to generate novel elastin-mimetic polypeptides, where the analogue L-**DOPA** servers as a viable substrate during *in vivo* protein synthesis. The residue-specific incorporation strategy permits global substitution of the analogue in response to the tyrosine codon at nearly twenty-one positions, representing an especially high degree of biosynthetic substitution. Under conditions favorable for hydrophobic assembly, the polymer undergoes a reversible sol-gel transition over a narrow temperature range with an inverse temperature profile characteristic of model elastin-like polypeptides. We envision that these novel elastin-mimetic polypeptides containing the unnatural amino acid DOPA will have advantages over previously designed polymers. The presence of the dihydroxy-phenylalanine group periodically along the polymer chain may serves as attractive sites for a versatile number of chemical strategies for cross-linking and generating conjugates to other molecules. Preliminary investigations, demonstrate the elastin-mimetic polypeptide is capable of forming metal-ligand (DOPA) complexes and that in concentrated polymer solutions the generation of these metal-ligand cross-links results in formation of a hydrogel. The cross-linking strategy utilized in this study is an attractive method for generating polymeric hydrogels under potentially biologically relevant conditions.

References

[1] Drury, J. L.; Mooney, D. J. *Biomaterials* **2003**, *24*, 4337-4351.

[2] Lee, K. Y.; Mooney, D. J. *Chemical Reviews* **2001**, *101*, 1869-1880.

[3] Wright, E. R.; Conticello, V. P. *Advanced Drug Delivery Reviews* **2002**, *54*, 1057-1073.

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

[5] Sallach, R. E.; Cui, W.; Wen, J.; Martinez, A.; Conticello, V. P.; Chaikof, E. L. *Biomaterials* **2009**, *30*, 409-422.

[6] McMillan, R. A.; Conticello, V. P. *Macromolecules* **2000**, *33*, 4809-4821.

[7] Nagapudi, K.; Brinkman, W. T.; Leisen, J. E.; Huang, L.; McMillan, R. A.; Apkarian, R. P.; Conticello, V. P.; Chaikof, E. L. *Macromolecules* **2002**, *35*, 1730-1737.

[8] Bellingham, C. M.; Lillie, M. A.; Gosline, J. M.; Wright, G. M.; Starcher, B. C.; Bailey, A. J.; Woodhouse, K. A.; Keeley, F. W. *Biopolymers* **2003**, *70*, 445-455.

[9] Girotti, A.; Reguera, J.; Rodríguez-Cabello, J. C.; Arias, F. J.; Alonso, M.; Ma Testera, A. *Journal of Materials Science: Materials in Medicine* **2004**, *15*, 479-484.

[10] Lee, J.; Macoscko, C. W.; Urry, D. W. *Biomacromolecules* **2001**, *2*, 170-179.

[11] Lim, D. W.; Nettles, D. L.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2003**, *9*, 222-230.

[12] Lim, D. W.; Nettles, D. L.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2007**, *8*, 1463-1470.

[13] Martino, M.; Tamburro, A. M. *Biopolymers* **2001**, *59*, 29-37.

[14] McMillan, R. A.; Lee, T. A. T.; Conticello, V. P. *Macromolecules* **1999**, *32*, 3643-3648.

[15] Nowatzki, P. J.; Tirrell, D. A. *Biomaterials* **2004**, *25*, 1261-1267.

[16] Trabbic-Carlson, K.; Setton, L. A.; Chilkoti, A. *Biomacromolecules* **2003**, *4*, 572-580.

[17] Vieth, S.; Bellingham, C. M.; Keeley, F. W.; Hodge, S. M.; Rousseau, D. *Biopolymers* **2007**, *85*, 199-206.

[18] Zio, K. D.; Tirrell, D. A. *Macromolecules* **2003**, *36*, 1553-1558.

[19] Rodgers, K. J.; Dean, R. T. *International Journal of Biochemistry and Cell Biology* **2000**, *32*, 945-955.

[20] Waite, J. H. *Biological Bulletin* **1992**, *183*, 178-184.

[21] Deming, T. J. *Current Opinion in Chemical Biology* **1999**, *3*, 100-105.

[22] Rzepecki, L. M.; Waite, J. H. In *Bioorganic Marine Chemistry*; Scheuer, P. J., Ed.; Springer: New York, 1991; Vol. 4, p 119-148.

[23] Waite, J. H. *Integrative and comparative biology* **2002**, *42*, 1172-1180.

[24] Land, E. J.; Ramsden, C. A.; Riley, P. A. *Methods in Enzymology* **2004**, *378*, 88-109.

[25] Waite, J. H. *International Journal of Adhesion and Adhesives* **1987**, *7*, 9-14.

[26] Berenbaum, M. B.; Williams, J. I.; Bhattacharjee, H. R.; Goldberg, I.; Swerdloff, M. D.; Salerno, A. J.; Unger, P. D. *Polymer Preprints (American Chemical Society, Division of Polymer Chemistry)* **1989**, *30*, 350-351.

[27] Huang, L.; McMillan, R. A.; Apkarian, R. P.; Pourdeyhimi, B.; Conticello, V. P.; Chaikof, E. L. *Macromolecules* **2000**, *33*, 2989-2997.

[28] Swerdloff, M. D.; Anderson, S. B.; Sedgwick, R. D.; Gabriel, M. K.; Brambilla, R. J.; Hindenlang, D. M.; Williams, J. I. *International Journal of Peptide and Protein Research* **1989**, *33*, 318-327.

[29] Yamada, K.; Chen, T.; Kumar, G.; Vesnovsky, O.; Topoleski, T.; Payne, G. F. *Biomacromolecules* **2000**, *1*, 252-258.

[30] Yamamoto, H. *Journal of Adhesion Science and Technology* **1987**, *1*, 177-183.

[31] Yamamoto, H. *journal of the Chemical Society, Perkin Transactions 1* **1987**, 613-618.

[32] Yamamoto, H. *Biotechnology and Genetic Engineering Reviews* **1996**, *13*, 133-165.

[33] Yamamoto, H.; Yamauchi, S.; Ohara, S. *Biomimetics* **1992**, *1*, 219-238.

[34] Yu, M.; Deming, T. J. *Macromolecules* **1998**, *31*, 4739-4745.

[35] Yu, M.; Hwang, J.; Deming, T. J. *Journal of the American Chemical Society* **1999**, *121*, 5825-5826.

[36] Liu, B.; Burdine, L.; Kodadeck, T. *Journal of the American Chemical Society* **2006**, *128*, 15228-15235.

[37] Alfonta, L.; Zhang, Z.; Uryu, S.; Loo, J. A.; Schultz, P. G. *Journal of the American Chemical Society* **2003**, *125*, 14662-14663.

[38] Umeda, A.; Thibodeaux, G. N.; Zhu, J.; Lee, Y.; Zhang, Z. J. *ChemBioChem* **2009**, *10*, 1302-1304.

[39] Anderson, J. R., Emory University, 2004.

[40] Lee, B. P.; Chao, C.-Y.; Motan, E.; Nunalee, F. N.; Shull, K.; Messersmith, P. B. *Macromolecules* **2006**, *39*, 1740-1748.

[41] Lee, B. P.; Dalsin, J. L.; Messersmith, P. B. *Biomacromolecules* **2002**, *3*, 1038-1047.

[42] Lee, B. P.; Huang, K.; Nunalee, F. N.; Shull, K. R.; Messersmith, P. B. *Journal of Biomaterials Science, Polymer Edition* **2004**, *15*, 449-464.

[43] Westwood, G.; Horton, T. N.; Wilker, J. J. *Macromolecules* **2007**, *40*, 3960-3964.

[44] Holten-Andersen, N.; Harrington, M. J.; Birkedal, H.; Lee, B. P.; Messersmith, P. B.; Lee, K. Y.; Waite, J. H. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 3-7.

[45] Harrington, M. J.; Masic, A.; Holten-Andersen, N.; Waite, J. H.; Fratzl, P. *Science* **2010**, *328*, 216-220.

[46] Holten-Andersen, N.; Mates, T. E.; Toprak, M. S.; Stucky, G. D.; Zok, F. W.; Waite, J. H. *Langmuir* **2009**, *25*, 3323-3326.

[47] Zeng, H.; Hwang, D. S.; Israelachvili, J. N.; Waite, J. H. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 12850-12853.

[48] Avdeef, A.; Sofen, S. R.; Bregante, T. L.; Raymond, K. N. *Journal of the American Chemical Society* **1978**, *100*, 5362-5370.

[49] Sever, M. J.; Weisser, J. T.; Monohan, J.; Srinivasan, S.; Wilker, J. J. *Angewandte Chemie International Edition* **2004**, *43*, 448-450.

[50] Taylor, S. W.; Chase, B.; Emptage, M. H.; Nelson, M. J.; Waite, J. H. *Inorganic Chemistry* **1996**, *35*, 7572-7577.

[51] Taylor, S. W.; Luther, G. W. I.; Waite, J. H. *Inorganic Chemistry* **1994**, *33*, 5819-5824.

[52] Lee, H.; Schere, N. F.; Messersmith, P. B. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 12999-13003.

[53] Carpenter, H. E., Emory University, 2008.

[54] Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. *Molecular Systems Biology* **2006**, *2*, 1-11.

[55] Budisa, N.; Steipe, B.; Demange, P.; Eckerskorn, C.; Kellermann, J.; Huber, R. *European Journal of Biochemistry* **1995**, *230*, 788-796.

[56] Sambrook, J.; Russell, D. W. *Molecular cloning: a laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

[57] Payne, S. C.; Patterson, M.; Conticello, V. P. In *Protein Engineering Handbook*; Lutz, S., Bornscheuer, U. T., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2009; Vol. 1, p 915-936.

[58] Kim, W.; George, A.; Evans, M.; Conticello, V. P. *ChemBioChem* **2004**, *5*, 928-936.

[59] Meyer, D. E.; Chilkoti, A. *Nature Biotechnology* **1999**, *17*, 1112-1115.

[60] Wishart, D. S.; Bigam, C. G.; Yao, J.; Abilgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. *Journal of Biomolecular NMR* **1995**, *6*, 135-140.

[61] Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2002**, *3*, 357-367.

[62] Kiick, K. L.; Tirrell, D. A. *Tetrahedron* **2000**, *56*, 9487-9483.

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

[64] Zhang, H.; Prasad, K. U.; Urry, D. W. *Journal of Protein Chemistry* **1989**, *8*, 173-182.

[65] Fancy, D. A.; Kodadeck, T. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 6020-6024.

[66] Urry, D. W.; Okamoto, K.; Harris, R. D.; Hendrix, C. F.; Long, M. M. *Biochemistry* **1976**, *15*, 4083-4089.

[67] McHale, M. K.; Setton, L. A.; Chilkoti, A. *Tissue Engineering* **2005**, *11*, 1768-1779.

[68] Sever, M. J.; Wilker, J. J. *Dalton Transactions* **2006**, 813-822.

Chapter 5

Site-specific, multi-site incorporation of a photo-cross-linkable amino acid analogue via a novel amber suppression strategy

Introduction

Protein-based polymers generated by recombinant-gene-expression methods are being developed to create 'designer materials' with tightly defined physical, chemical, and biological properties. Protein engineering affords synthetic matrices with controlled sequence as well as defined features and tunable compositions, organization, mechanics, and biological responses. These approaches have been exploited to create tailor-made hydrogels, stimuli responsive polymers and materials with controlled biorecognition, cross-linking, degradation, structure and mechanical properties[1]. While the primary level control of polypeptide structure makes such strategies advantageous over conventional synthetic routes, the chemical diversity is limited to only 20 canonical amino acids. Native proteins, however, require additional chemical groups to arrange into the complex architectures responsible for the materials performance of tissues and to perform the functions needed for life. Expanded chemical functionality is generally accomplished through covalent post-translational modifications of the natural amino acid side chains, such as acetylation, methylation, and phosphorylation, to name a few. In an effort to recreate the chemical complexity of native proteins, chemists and biochemists have investigated manipulation of expression host translational machinery to recognize unnatural amino acids during *in vivo* synthesis of polypeptides. Unnatural amino acids are characterized by unique functional groups that impart novel chemical properties to the synthetic polypeptides[2]. Strategies for unnatural amino acid incorporation during *in vivo* protein synthesis may be useful for generating polymer mimetics in high yields, which emulate or expand upon the functions of their natural counterparts.

Residue-specific incorporation has been utilized for the introduction of a variety of structurally similar unnatural amino acids in response to the natural amino acid codon[3]. This biosynthetic incorporation method has been utilized for the global substitution of non-canonical amino acids to generate polypeptides with improved conformational stability, in the case of fluorinated analogues[4-6], to control peptide self-assembly[7,8], to investigate protein-protein interactions, to probe protein structure[9,10], and to introduce photo- and chemically reactive moieties into proteins. For example, the residue-specific incorporation strategy has been utilized for the construction of polymer mimetics of the extracellular matrix (ECM) proteins like collagen. Like many fibrous proteins of the ECM, the amino acid sequence of native collagen is characterized by a repeating oligopeptide motif. Specifically, collagen is characterized by the tripeptide consensus motif Pro-Yaa-Gly where Yaa represents a genetically incorporated proline residue that is post-translationally modified to generate hydroxyproline. Under selective pressure, Buechter and co-workers found that the unnatural amino acid analogue 4-hydroxyproline (Hyp), could be used to replace the proline residues in a Type I collagen-like domain, which exhibited thermal stability comparable to fragments of native Type I collagen[11]. In addition, the native elastin network found in the ECM is generated through post-translational modifications that play an important role in its elastomeric properties. Specifically, oxidative deamination of lysyl residues in the soluble precursor tropoelastin results in covalent cross-linking of the polymer chains. Cross-linking in native elastin is extensive and significantly influences the biophysical properties of the protein. Efforts to introduce covalent cross-links into elastin-mimetic polypeptides have utilized chemical cross-linking strategies that take advantage of genetically encoded lysine residues. Alternatively, Chapters 3 and 4 of this dissertation have

described the incorporation of unnatural amino acid residues within elastin-mimetic polypeptides, which have the potential to form cross-linked networks or be further derivatized under physiologically relevant conditions.

Engineering of these ECM polymer mimics containing unnatural amino acids illustrates the limitations of the residue-specific incorporation method. Specifically, the method is not site-specific and is limited to use of structurally similar unnatural amino acids that represent only a subset of the potentially useful chemical functionalities. With regard to the synthesis of a collagen-mimetic polypeptide, the indiscriminate insertion of Hyp at proline codons does not allow for production of the naturally occurring Pro-Hyp-Gly repeating peptide. Furthermore, the set of unnatural amino acids available for introducing expanded chemical functionality is limited to structurally similar analogues that act as viable substrates for wild type or mutant aminoacyl-tRNA synthetase. Thus, the residue-specific method is not suitable for engineering of proteins where site-specific incorporation is desired or the unnatural amino acid analogue under investigation is structurally dissimilar. A site-specific method for genetic incorporation of non-canonical amino acids directly into proteins during *in vivo* translation, however, represents an alternative method for installing a wide range of amino acid analogues.

The site-specific incorporation of unnatural amino acids during *in vivo* protein synthesis requires the addition of new components to the host strain biosynthetic machinery. Specifically, these include the unnatural amino acid target, a unique codon, a corresponding tRNA, and a cognate aminoacyl-tRNA synthetase. Work by Schultz and coworkers has utilized selection schemes[12] to generate heterologous pairs derived from amber suppressor tRNAs and evolved mutant synthetases, which function orthogonally to incorporate the analogue in response to the

termination codon using amber suppression[13]. The amber suppression method permits site-specific substitution of the non-canonical amino acid at any site along the polypeptide chain. The method has been successfully used to incorporate unnatural amino acid analogues, which represent commonly found post-translational modifications as well as structurally dissimilar derivatives, into proteins in *E. coli*[14,15]. Suppression of the amber codon, however, is characterized by limited efficiency[16]. Thus, the method is not ideal for multi-site incorporation of analogues and imposes limits on the degree of polypeptide derivatization. The low efficiency of amber suppression is a result of competition by the endogenous *E. coli* release factor (i.e. RF1), which causes release of the premature peptide from the ribosome. Altering the rate of RF1-mediated translation termination may enhance amber suppression and thus, permit incorporation of multiple non-canonical amino acids within a single polypeptide. A recent publication has demonstrated the incorporation of an unnatural amino acid analogue of lysine characterized by acetylation at multiple sites within a green fluorescent protein encoding three UAG codons[17]. The authors overexpressed the C-terminal domain of a small ribosomal protein L11, which was previously implicated in RF1-mediated peptide release[18,19]. While the method permitted synthesis of a full-length GFP protein, protein yields remained low. In this strategy, the authors cloned both the target protein gene and the orthogonal tRNA/synthetase pair on the same expression plasmid and subsequently cotransformed the recombinant molecules with a plasmid expressing the L11C protein in the expression host strain. In order to assess the ability of the strategy to promote amber suppression at an increased number of sites within the same mRNA transcript, the target protein expression plasmid would require re-cloning. Furthermore, the use of a new heterologous tRNA/synthetase pair and unnatural amino acid substrate would also

require a time-intensive number of cloning steps. Therefore, we wished to develop a strategy that would facilitate the synthesis of the target protein gene with convenient selection of number of amber suppression sites and permit the use of a variety of orthogonal tRNA/synthetase pairs.

We propose a simple strategy for increased amber suppression that employs a mutant *E. coli* expression strain with altered RF1 activity, which was previously isolated by Rydén-Aulin and coworkers. The authors originally isolated a mutant form of RF1, *prfA1*, which caused a modified phenotype where growth at 42 °C was temperature sensitive and resulted in increased readthrough of stopcodons UAA and UAG[20,21]. In a later publication, the authors isolated a mutant that compensated for the Ts phenotype caused by *prfA1*. The newly isolated strain was characterized by a mutation *rpsD101* in the ribosome protein S4, which is part of the domain with which RF1 and RF2 interact when binding to the ribosome during translation termination[22]. Previous strains having a mutation in the *rpsD* gene show a compensating phenotype for increased stopcodon readthrough, but the interaction between the *rpsD101* and *prfA1* mutation appears to be unique. Specifically, instead of suppressing the increased misreading of UAG caused by *prfA1*, the *rpsD101* mutation actually enhances it. We hypothesize that the increased readthrough of stopcodon UAG in strain MRA317 should permit multiple amber suppression events during translation of a single mRNA transcript. In this study, we were particularly interested in the site-specific, multi-site incorporation of the tyrosine analogue, *para*-benzoyl-L-phenylalanine (*p*Bpa) (Figure 1).

Benzophenones represent a class of photoactivatable probes that have been utilized extensively for the study of protein-protein interactions for biological systems both *in vit*ro and *in vivo*[23]. The distinct chemical and biochemical advantages of benzophenones make

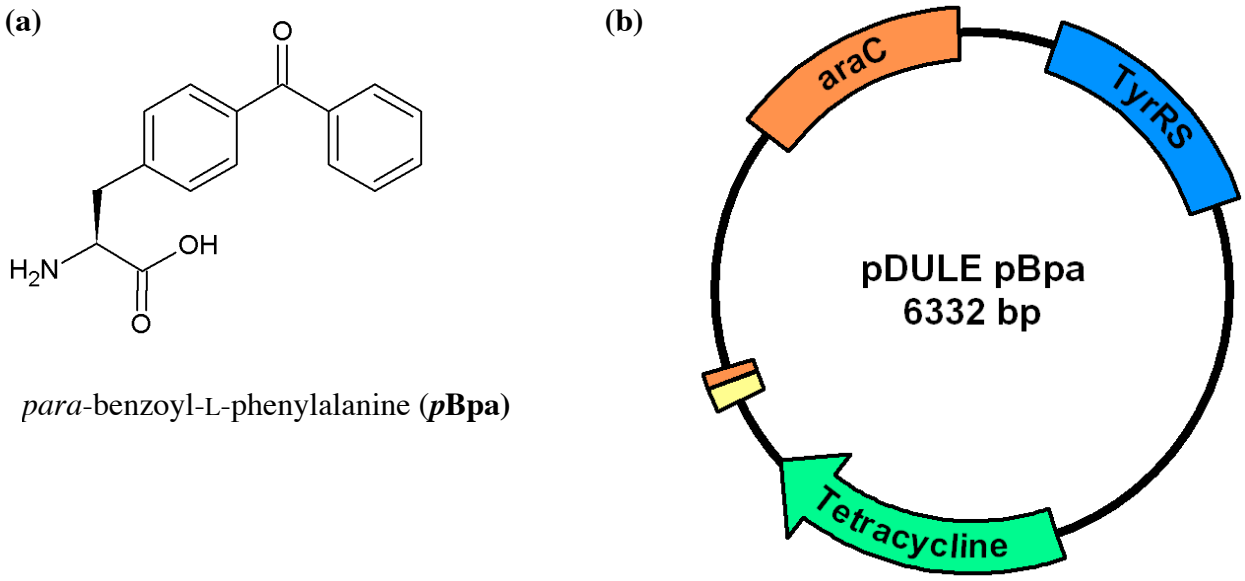
them attractive photo-initiated cross-linkers for incorporation into the protein polymer backbone[23,24]. First, the benzophenone class of photoprobes is chemically more stable than other photoactivatable compounds, including diazo esters, aryl azides, and diazirines. Also, the benzophenones can be easily manipulated under ambient light and can be activated at wavelengths of 350-360 nm, avoiding protein-damage[25,26]. Third, the irradiated benzophenones give rise to a triplet biradical, which reacts preferentially with the unreactive C-H bonds of protein chains, even in the presence of solvent water and bulk nucleophiles[27]. Furthermore, benzophenones do not photodissociate and their photoactivated triplet state readily relaxes in the absence of an unactivated C-H bond with which to react[24]. This reversible excitation allows for repeated excitation by UV light without damage to protein and thus, higher cross-linking yields.

The photo-cross-linking chemistry of benzophenones has been widely exploited to cross-link small molecules or peptides to proteins[27]. Until recently, however, no simple method existed for the site-specific incorporation of benzophenones into larger proteins inaccessible by solid-phase synthesis. A publication in *PNAS* by the Schultz group demonstrated *in vivo* incorporation of *p*Bpa through the use of an orthogonal aminoacyl-tRNA synthetase/tRNA pair in response to the amber codon UAG[24]. The method utilized an engineered mutant suppressor $tRNA^{Tyr}$ from *Methanococcus jannaschii* ($mutRNA^{Tyr}$), which is aminoacylated by the cognate *M. jannaschii* TyrRS (MjTyrRS) more efficiently than any endogenous *E. coli* synthetases[12,28,29]. The *M. jannaschii* pair efficiently suppresses the amber codon by incorporating tyrosine at the UAG codon. Since the *M. jannaschii* synthetase does not aminoacylate any endogenous *E. coli* tRNAs with tyrosine, only the mutant tyrosine amber suppressor, the specificity of the orthogonal TyrRS can be altered for a desired unnatural amino acid. Schultz and coworkers demonstrated selection

of a mutant MjTyrRS that aminoacylates the mutRNA^{Tyr} with the unnatural amino acid *p*Bpa with high fidelity and efficiency[24]. Schultz and coworkers have also demonstrated *in vitro* photo-cross-linking with incorporation of the *p*Bpa site-specifically *in vivo* by cross-linking glutathione S-transferase from *Schistosoma japonica*. In a later publication, Schultz and Chin demonstrated the *in vivo* photo-cross-linking of interacting glutathione S-transferase at sites of *p*Bpa incorporation[30]. The practical utility of organisms with an expanded genetic code is dependent on the efficiency and fidelity of analogue incorporation. Improvements to the components of the expression system have been demonstrated and include optimization of the orthogonal tRNA/synthetase expression vector. Farrell and coworkers demonstrated improved yields of *p*Bpa containing polypeptides through development a plasmid coexpressing the *M. jannaschii* amber suppressor tRNA and mutant synthetase[31]. Based on these results the pDULE-Bpa (Figure 1) plasmid will be utilized in this study.

In summary, we propose a novel amber suppression strategy that utilizes an amber suppressor tRNA/synthetase pair, the unnatural amino acid *p*Bpa, and the mutant *E. coli* expression strain MRA317. The orthogonal pair will be expressed from the plasmid pDULE-Bpa and direct the incorporation of *p*Bpa in response to the UAG codon. Expression of the artificial coding sequences, which contain increasing amber codons, will be placed under control of an IPTG inducible promoter. Expression will be carried out in a mutant *E. coli* strain, characterized by increased readthrough of the stopcodon UAG. We hypothesize that the orthogonal tRNA/synthetase pair will function in the host strain to incorporate the unnatural amino acid site-specifically at multiple sites within the growing polypeptide chain. Furthermore, the proposed scheme will utilize a two-plasmid expression system with coexpression of a plasmid encoding

the target polypeptide and a second plasmid encoding the orthogonal tRNA/synthetase pair. In this way, a variety of target proteins can be synthesized with a wide array of unnatural amino acid substrates through simple transformation of the plasmids.



para-benzoyl-L-phenylalanine (***p*Bpa**)

Figure 1. (a) Structure of *p*Bpa and (b) a schematic of the pDULE-Bpa plasmid. The *p*Bpa-specific aminoacyl-tRNA synthetase gene (blue) is located downstream of the *lpp* promoter and before the *rrnB* terminator. The amber suppressor tRNA is shown in yellow.

Experimental Methods

Materials

All chemical reagents were purchased from Fisher Scientific, Inc. (Pittsburgh, PA), Sigma-Aldrich Corporation (St. Louis, MO), or VWR International LLC (Radnor, PA) unless otherwise noted. Tyrosine analogue, *para*-benzoyl-L-phenylalanine (***p*Bpa**), was purchased from Bachem Americas, Inc. (Torrence, CA). Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). Restriction endonucleases, antartic phosphatase, T4 DN ligase, T4 polynucleotide kinase, DNA polymerase I large (Klenow) fragment, and deoxynucleotide solution mix (dNTPs) were purchased from New England Biolabs, Inc. (Beverly, MA). Platinum® *Pfx* DNA polymerase and plasmid pZErO®-1 were obtained from Invitrogen Corp. (Carlsbad, CA).

Plasmid pQE-80L was purchased from QIAGEN Inc.-USA (Valencia, CA) while all plasmids designated pIL were synthesized by I-Lin Wu, a graduate student in the Conticello lab at Emory University. Plasmid pDULE-Bpa was generously provided by Dr. Ryan Mehl of Franklin and Marshall College (Lancaster, PA)[31]. Plasmid DNA preparation and purification were performed using the QIAfilter Plasmid Maxi Kit, QIAprep Spin Miniprep Kit, and QIAquick PCR purification Kit (QIAGEN Inc.-USA, Valencia, CA) and the DNA Clean & Concentrator™ and Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA). Synthetic oligonucleotides were purchased from Sigma-Genosys (a division of Sigma-Aldrich Corporation, St. Louis, MO) and were used as received. The codon optimized variant of superfolder green fluorescent protein (GFP) was synthesized by DNA2.0, Inc. (Menlo Park, CA). The *E. coli* strains TOP10F' and DH10B™ were obtained from Invitrogen Corp. (Carlsbad, CA).

The *E. coli* strain MRA317 was generously provided by Professor Monica Rydén-Aulin from Stockholm University (Stockholm, Sweden).

General Methods

The molecular biology techniques utilized here, including expression cloning polymerase chain reaction, gel electrophoresis, and growth and induction of bacterial cultures were adapted from a standard molecular cloning manual[32] or the protocol supplied by manufacturer, unless otherwise noted. Reagents for the manipulation of DNA, bacteria, and recombinant proteins were sterilized by either autoclave or passage through a syringe filter (0.2 μm cellulose membrane(or vacuum filter unit (standard polyethersulfone (PES) membrane) available from VWR International, LLC (Radnor, PA). Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. The concentration of DNA solutions was measured by an Ultrospec 3000 UV/Visible Spectrophotometer (Pharmacia Biotech, Cambridge, UK). Site-directed mutagenesis was performed using Stratagene's (La Jolla, CA) Quick-Change mutagenesis protocol from gene-specific oligonucleotide primers. Polymerase chain reaction (PCR) was carried out using a MJ Mini™ Gradient Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules, CA). Over the course of these studies automated DNA sequencing was performed at either Beckman Coulter Genomics (Danvers, MA) on a Perkin-Elmer ABI Prism 377 DNA sequencer or GENEWIZ, Inc. (South Plainfield, NJ) on an Applied Biosystems 3730xl DNA analyzer. Agarose gel electrophoresis images were captured using a Kodak Gel Logic 112 imaging system from Carestream Health, Inc. (Rochester, NY). Protein electrophoresis was performed using the Mini-PROTEAN® 3 Cell electrophoresis system from Bio-Rad Laboratories, Inc. (Hercules, CA) and the Perfect Protein™ Markers, 15-150 kDa from

Novagen® (a brand of EMD Chemicals Inc., Gibbstown, NJ). Protein expression was detected by Western blot using the His•tag® Monoclonal Antibody and the His•tag® AP Western Reagents Kit, which includes the Goat Anti-Mouse IgG AP Conjugate (H+L) and AP Detection Reagents for development, both from Calbiochem® (a division of EMD Chemicals, Inc. Gibbstown, NJ). Proteins were purified by metal-affinity chromatography using the cobalt-charged TALON® Metal Affinity Resin from Clontech Laboratories (Mountainview, CA).

Construction of plasmids

Elastin-UAG expression plasmids

The **elastin-UAG** multimers, which encode a highly repetitive, elastin-mimetic polymer, were constructed using a previously reported concatemerization strategy[33]. The procedure is briefly summarized below. DNA oligonucleotide primers encoding the sense and anti-sense strand of an elastin-UAG monomer, **elastin-UAG-F** (5'-AGCTTGAAGACGTTCCAGG TGTTGGCGTACCGGGTGTAGGCGTTCCGGGTTAGGGTGTTCCAGGCGTAGGTGTACC GGGTGTAGGTGTTCCAAGAGACGG-3') and **elastin-UAG-R** (5'AGCTTGAAGACGTT CCAGGTGTTGGCGTACCGGGTGTAGGCGTTCCGGGTTAGGGTGTTCCAGGCGTAGGT GTACCG GGTGTAGGTGTTCCAAGAGACGG-3') were chemically synthesized and annealed to produce duplex monomer DNA. The DNA monomer was purified by agarose gel electrophoresis and phosphate groups were added to the 5'-end of the DNA duplex with T4 polynucleotide kinase. The duplex monomer DNA was ligated into the *Bam*H I/*Hin*d III-digested pZErO®-1 cloning plasmid. The ligation mixture was used to transform electrocompetent cells of *E. coli* strain TOP10F' and transformants were selected on low salt LB agar media supplemented with 1 mM IPTG and Zeocin™ (50 μg/ml) via incubation at 37 °C for 12-14 h. Recombinant

plasmid DNA was isolated from single colonies using the QIAprep Spin Miniprep Kit and screened via double digest with *Bam*H I and *Hin*d III. Plasmid DNA containing an insert size concomitant with the monomer gene was then analyzed by automated DNA sequencing using the M13 reverse and M13 (-20) forward primers (Appendix 1).

DNA cassette concatemerization was performed by first preparing a large amount of elastin-UAG monomer plasmid DNA using the QIAfilter Plasmid Maxi Kit. The elastin-UAG DNA monomer was liberated from the cloning plasmid via sequential digestion with *Bbs* I and *Bsm*B I and purified via agarose gel electrophoresis (2% NuSieve®GTG® agarose; 1X TBE buffer). Self-ligation of the DNA monomer in a head-to-tail fashion was conducted by incubating with T4 DNA ligase and T4 DNA ligase buffer with ATP at 16 °C for 16 h. The library of DNA concatemers was fractionated via agarose gel electrophoresis (1% agarose, 0.5X TBE buffer). The region of the gel corresponding to the desired multimer size range (~1000 to 2000 bp) was excised from the gel and purified using the Zymoclean™ Gel DNA Recovery Kit. The purified DNA concatemers were then ligated back into the *Bbs* I-digested and phosphorylated elastin-UAG monomer acceptor plasmid. The ligation mixture was used to transform electrocompetent cells of *E. coli* strain TOP10F' and transformants were selected on low salt LB agar media supplemented with 1 mM IPTG and Zeocin™ (50 μg/mL). The recombinant plasmids were screened via double digestion with *Bam*H I and *Hin*d III and analyzed by agarose gel electrophoresis (1% agarose, 0.5X TBE buffer) for the presence of multimer inserts. From the pool of transformants, two colonies were isolated containing recombinant plasmids with concatemer inserts of approximately 1200 and 2000 bp. The sequence of the recombinant

plasmids was confirmed by automated DNA sequencing analysis using the M13 reverse and M13 (-20) forward primers.

An expression plasmid was prepared via insertion of a modified polylinker sequence into the commercially available pQE-80L to generate plasmid **pIL2**. The modified polylinker contained recognition sites for type IIs restriction endonucleases essential for seamless cloning of the elastin-UAG concatemers. For subsequent cloning of the elastin-UAG concatemers, it was necessary to remove a *Bsa* I restriction endonuclease recognition site from the modified expression plasmid. Site-directed mutagenesis was utilized to introduce a silent mutation into the sequence of the plasmid and produce the plasmid **pIL5**. The modified expression plasmid **pIL5** contained a copy of the overproducing repressor allele $lacI^q$ to ensure tight control of the basal level of transcription prior to induction with IPTG.

The elastin-mimetic DNA concatemers were then cloned into the modified pQE-80L expression vector **pIL5**. The elastin-UAG concatemers were removed from the pZErO®-1 cloning plasmids via sequential digestion with *Bbs* I and *Bsm*B I to isolate the multimer gene. The **pIL5** plasmid was digested with *Bsa* I to generate the complementary sticky ends necessary for seamless cloning of the concatemers into the adaptor sequence. Prior to ligation, the linearized plasmid DNA was also dephosphorylated with antartic phosphatase to reduce the incidence of false positive ligations. The ligation mixtures were then used to transform electrocompetent cells of TOP10F' and the recovery mixture was plated onto LB agar media supplemented with ampicillin (100 μg/mL). Plasmid DNA was isolated from single colony transformants and screened via double digestion with *Eco*R I and *Hin*d III for the presence of the elastin-UAG concatemer inserts. The sequence of the expression plasmids were confirmed by

automated DNA sequencing analysis using primers designed to bind the pQE-80L plasmid upstream and downstream of the adaptor sequences (Appendix 1). The correctly sequenced plasmids **pIL43.5** and **pIL43.11** include the pQE-80L plasmid with insertion of concatemers approximately 1200 and 2000 bp in length, respectively. The adaptor sequence was designed to include a decahistidine tag at the C-terminus of the elastin-UAG polypeptides to facilitate visualization of proteins by Western blot and purification of the proteins by metal affinity chromatography. The placement of the decahistidine tag at the C-terminus should ensure the identification and purification full-length protein products selectively.

sfGFP expression plasmid

A synthetic gene encoding a superfolder green fluorescent protein (sfGFP)[34] variant with three amber codons inserted at sites between amino acid residues 157-158, 172-173, and 194-195 was chemically synthesized. The DNA was digested with *Bam*H I and *Eco*R I and cloned into compatible sites in the modified pQE-80L expression plasmid **pIL5**. In this way, the sfGFP expression cassette is placed upstream of an in-frame decahistidine tag, which will facilitate visualization of proteins by Western blot and purification of the proteins by metal affinity chromatography. The placement of the decahistidine tag at the C-terminus should ensure the identification and purification full-length protein products selectively.

Bacterial growth and expression

Chemically competent cells of *E. coli* strain MRA317 [*prfA1*, *rpsD101*][35] were prepared. Cells were transformed with the plasmids, **pIL80**, **pIL43.5** and **pIL43.11**, encoding the target proteins and selected on LB agar media supplemented with ampicillin (100 μg/mL) via incubation at 37 °C for 12-14 h. Fresh chemically competent cells were prepared from single

colonies and transformed with the plasmid pDULE-Bpa. The transformation mixture was plated on LB agar media supplemented with ampicillin and tetracycline (25 μg/mL) and plates were incubated at 37 °C for 12-14 h. In this way, the expression strains employed for these studies were constructed.

Single colonies of each expression strain were used to inoculate 20 mL LB media supplemented with ampicillin (100 μg/mL) and tetracycline (25 μg/mL). The cultures were then incubated overnight at 37 °C with agitation at 225 rpm. One liter of LB media was prepared for each expression strain in 2-2.8 L flasks (500 mL each). Each flask was supplemented with antibiotics, ampicillin (50 μg/mL) and tetracycline (2.5 μg/mL), at reduced concentrations to maximize protein yield. Each liter of fresh LB media was inoculated with 20 mL (2% final culture volume) of overnight cultures and incubated at 37 °C with shaking at 225 rpm until the optical density, as measured at wavelength 600 nm, reached 0.8 to 1.0 absorbance units. A 200 mM stock solution of the *para*-benzoyl-L-phenylalanine (***p*Bpa**) analogue was prepared fresh in phosphate buffered saline (PBS) and titrated with 1 M NaOH until soluble. The unnatural amino acid stock solution was then added to the cultures to a final concentration of 2 mM. Expression of the sfGFP and elastin-mimetic proteins was induced by adding IPTG to a final concentration of 1 mM to the cultures. The cultures were incubated at 37 °C with agitation at 225 rpm for 20 h after induction. The cells were then harvested by centrifugation at 4000 x *g* and 4 °C for 20 min. The cell pellet was suspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and stored at -80 °C.

TALON metal-affinity column purification

Three freeze (-80 °C) /thaw (25 °C) cycles were employed for initial cell fracture. Lysozyme from chicken egg white was added to the cell lysate at a final concentration of 1 mg/mL (from a 10 mg/mL stock solution) along with 1X Protease Inhibitor Cocktail Set V, EDTA-free from Calbiochem® (a brand of EMD Chemicals Inc., Gibbstown, NJ). The mixture was incubated at room temperature for 30 min with agitation. Benzonase® nuclease and 1 M $MgCl_2$ were added to the cell suspension for a final concentration of 25 units/mL and 1 mM, respectively. The solutions were incubated at 4 °C for 36 h. The cell suspension was centrifuged at 20,000 x *g* for 40 min at 4 °C.

The supernatant and pellet were then analyzed by western blot analysis, using the anti-his•tag monoclonal antibody, to determine the location of the target protein. For purification of the soluble sfGFP and the elastin-mimetic protein encoded by **pIL43.5**, the supernatant was loaded directly onto a column containing cobalt-charged TALON® metal affinity resin (3 mL) and washed with lysis buffer (30 mL) containing 20 mM imidazole. The target protein was eluted with lysis buffer (20 mL) containing 250 mM imidazole. The eluate was dialyzed at 4 °C using the SnakeSkin® Pleated Dialysis Tubing with a molecular weight cut-off of 10 kDa (Pierce Protein Research Products, Thermo Fisher Scientific, In., Rockford, IL) against distilled, deionized water (5 x 4 L). The dialysate of sfGFP was concentrated by ultracentrifugation using the Amicon Ultra-15 Centrifugal Filter Units (NMWL: 10 kDa) from Millipore (Burlington, MA) to a final volume of 1 mL. Protein concentration was determined by the Bradford assay using the Bio-Rad Protein Assay (Hercules, CA) by comparison with a BSA standard curve.

Following dialysis, the elastin-mimetic protein was lyophilized to produce a white, spongy solid and the yield was measured by dry weight.

The insoluble elastin-mimetic protein encoded by **pIL43.11** was purified from the cell pellet under denaturing conditions. The pellet was suspended in denaturing lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, 6 M urea, pH 7.0) and incubated at 4 °C with shaking overnight. The cell suspension was then centrifuged at 4 °C, 20,000 x *g* for 40 min. The cell lysate was then loaded onto a column containing cobalt-charged TALON® metal affinity resin (3 mL) and washed with lysis buffer (30 mL) containing 20 mM imidazole and 6 M urea. The target protein was eluted with lysis buffer (20 mL) containing 250 mM imidazole and 6 M urea. The elution fraction was dialyzed against a decreasing urea step gradient (6 M to 1 M) and further, against distilled, deionized water (5x 4 L) at 4 °C. The dialysate was lyophilized to produce a white, spongy solid and the yield was measured by dry weight.

FACS analysis

Chemically competent cells of *E. coli* strain MRA317 were transformed with the plasmids **pIL80** (encoding the sfGFP variant with insertion of three-UAG codons) and pDULE-Bpa (encoding the *M. jannaschii* orthogonal amber suppressor tRNA and mutant TyrRS synthetase pair). Transformants were identified by antibiotic selection on LB agar media with ampicillin (100 μg/mL) and tetracycline (25 μg/mL) at 37 °C. Single colonies were used to inoculate LB media (5 mL) supplemented with ampicillin and tetracycline and grown overnight at 37 °C in a rotating drum. Cells were subcultured in fresh LB media (5 mL) supplemented with ampicillin (50 μg/mL) and tetracycline (2.5 μg/mL) by inoculation with 100 μL overnight culture. The tubes were incubated at 37 °C in a rotating drum until cultures reach log growth

phase as measured by an $OD_{600\ nm}$ of approximately 0.8 to 1.0. Aliquots (1 mL) of the cultures were centrifuged at 4000 x *g* and 4 °C for 10 min. The supernatant was discarded and cells were gently suspended in 1 mL of phosphate-buffered saline pH 7.4. Flow cytometry was performed using a LSRII (Beckton Dickinson) equipped with a 100mW solid-state laser emitting at 488nm for the excitation of sfGFP, 505nm LP dichroic mirror and 530/30 bandpass filter. Forward scatter (FSC), sideward scatter (SSC) and green fluorescence (Detector E) were acquired by FACSDiva software. The specific instrumental gain settings for these measurements were as follows: FSC=250, SSC=300, F1=302). The maximum of each fluorescence histogram (number of events as a function of fluorescence) was scaled to 10000 to facilitate comparison of the histograms of sfGFP variants. Data was analyzed using FlowJo (TreeStar.com).

Thermolysin digestion

Purified elastin derivatives were dissolved in sterile water at a concentration of approximately 1 mg/mL. Dithiothreitol (DTT) was added as final concentration 10 mM and the mixture was incubated at 100 °C for 30 min to denature the protein. After the reaction mixture was cooled to room temperature, thermolysin was added to the denatured protein solution to a concentration of 1 part in 50 (w/w) with respect to protein. Since the *p*Bpa-containing elastin-mimetic polypeptides formed a turbid solution at 37 °C, the thermolysin digestion was carried out at 25 °C for 12 h. The products from the proteolysis reaction were passed through a C18 spin column (Thermo scientific) to remove salts and thermolysin. The peptide fragments were eluted using a 70:30 acetonitrile: water mixture. The solution was employed directly for mass spectrometric analysis.

Mass spectrometry

ESI spectra were acquired by LTQ-FT mass spectrometer (ThermoElectron, San Jose, CA) in a positive mode (200-2000 Da) using spray voltage of 4kV, capillary voltage 41V, sheath gas flow rate 20 arbutary units, tube lens voltage was 215V, and AGC setting was 5e+05. Molar masses of full-length elastin-mimetic polypeptides were determined by MALDI-TOF MS on an Applied Biosystems Voyager System STR mass spectrometer in the positive linear mode. Ferulic acid was used as the matrix at a concentration of 20 mg/mL in a mixture of 75 % acetonitrile: 0.2% formic acid. The protein solution (1 mg/mL in distilled water) was mixed with the matrix solution in a ratio of 1:10 and dried under air. Bovine serum albumin (BSA) was used as a standard for external calibration. Tandem MALDI-TOF/TOF MS experiments were conducted by Professor Gianluca Giorgi of the Department of Chemistry at the University of Siena (Siena, Italy).

Results and Discussion

Genes encoding three target artificial proteins were constructed and cloned into the modified pQE-80L plasmid, where each genetic construct contains an increasing number of UAG codons within the DNA coding sequence (Figure 2 and Scheme 1). Incorporation of the tyrosine analogue, *p*Bpa, during *in vivo* protein synthesis employed the coexpression of the plasmid pDULE-Bpa[31] (Figure 1). The pDULE-Bpa plasmid expresses an amber suppressor tRNA and mutant tyrosyl-tRNA synthetase both derived from *Methanococcus jannaschii*. The orthogonal pair has permitted site-specific, *in vivo* incorporation of the benzophenone derivative protein synthesis in *E. coli* in response to the amber codon TAG[24]. The *p*Bpa-specific aminoacyl-tRNA synthetase and $tRNA_{CUA}$ gene are both located downstream of the constitutively active *lpp* promoter. The plasmid contains a tetracycline resistance marker and a p15A origin, which makes it compatible with the modified pQE-80L plasmid comprising the expression clones. The incorporation of the benzophenone analogue was assayed in the modified *E. coli* strain MRA317[35], which is characterized by increased readthrough of the stopcodon UAG. The modified host strain was employed to facilitate amber suppression at multiple sites during translation of a single mRNA transcript and thus, increase analogue incorporation efficiency. The plasmids and strains utilized in this study are described in Table 1.

Fluorescence activated cell sorting (FACS) was used to identify the expression of the full-length, folded sfGFP. Initially, we examined the ability of the modified host strain to promote increased amber suppression and multi-site incorporation of *p*Bpa at three sites within a superfolder GFP variant. In addition, a positive control experiment was performed by expressing the wild-type sfGFP gene in the MRA317 strain under similar expression conditions. Cultures of

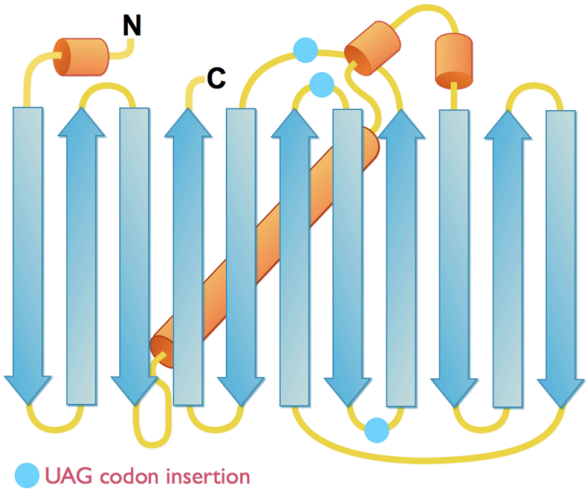


Figure 2. Secondary structure of superfolder GFP depicting the sites of UAG codon insertion (between 157-158, 172-173, and 194-195). The sites of insertion were selected such that incorporation of the unnatural amino acid analogue would not disrupt proper folding of the sfGFP variant. The structure is based on the PDB entry 2B3P.

Scheme 1. Amino acid sequence of **elastin-UAG** polypeptides, where the unnatural amino acid *p*Bpa is incorporated in response to the stopcodon UAG (indicated by a star). The elastin-mimetic polypeptides are designated **elastin-Bpa$_{12}$** (n = 12) and **elastin-Bpa$_{22}$** (n = 22).

$$M - [X] - VPGVGGSD_4KGH_{10}$$

$$[X] = [(VPGVG)_2VPG^*G(VPGVG)_2]_n$$

Table 1. List of recombinant plasmids and expression strains utilized in the expression of **sfGFP-Bpa** and **elastin-Bpa** polymers.

Recombinant Plasmid Name	**Plasmid/Insert**	**Comments**
pIL2	pQE-80L/**pQE Adaptor**	Amp^R
pIL5	pIL5 with removal of internal *Bsa* I	
pIL7	pIL5/wild-type superfolder GFP (sfGFP)	
pIL80	pIL5/**sfGFP** variant with insertion of three-UAG codons	
pIL43.5	pIL5/**elastin-UAG$_{12mer}$** concatemer	
pIL43.1	pIL5/**elastin-UAG$_{21mer}$** concatemer	
pDULE-Bpa	*M. jannaschii* orthogonal tRNA/synthetase pair	Tc^R; Reference[31]
Expression Strain Name	**Genotype**	**Comments**
DH10B		Invitrogen
MRA317	*prfA1*, *rpsD101*	Reference [35]

MRA317 cells transformed with **pIL80** and the pDULE-Bpa plasmid and supplemented with unnatural amino acid *p*Bpa showed fluorescence above the negative control expressions without supplementation of the analogue (Figure 3). The fluorescence of cells expressing the wild-type protein was significantly higher than the population of experimental cells. The decreased fluorescence with incorporation of *p*Bpa may indicate improper folding of the newly synthesized sfGFP variant.

Electrophoretic and western blot analysis of whole-cell lysates derived from these expression cultures indicate that the *p*Bpa analogue supported a detectable level of expression of the full-length, target protein with respect to the negative control culture, which was deficient in the analogue (Figure 4). Protein expression in rich media lacking the non-canonical amino acid is nearly undetectable by western blot analysis. This result supports that the amber suppressor tRNA/synthetase pair functions orthogonally and thus, the mutant synthetase is unable to charge the suppressor tRNA with the canonical amino acid tyrosine. A second control expression was performed using a conventional *E. coli* host strain DH10B with unaltered ribosomal activity and release factor recognition. Expression of the full-length sfGFP in rich media supplemented with tyrosine or the analogue was not observed in the unmodified host strain (data not shown). Thus, the modified host strain was necessary for expression of the full-length sfGFP variant from cultures supplemented with *p*Bpa.

Finally, the superfolder-GFP was purified to homogeneity by metal-affinity chromatography with an unoptimized yield of approximately 1 mg/L. This yield is comparable to other amber suppression incorporation strategies. The yield of the purified sfGFP containing the analogue was lower than expected and significantly lower than yields of wild-type sfGFP under

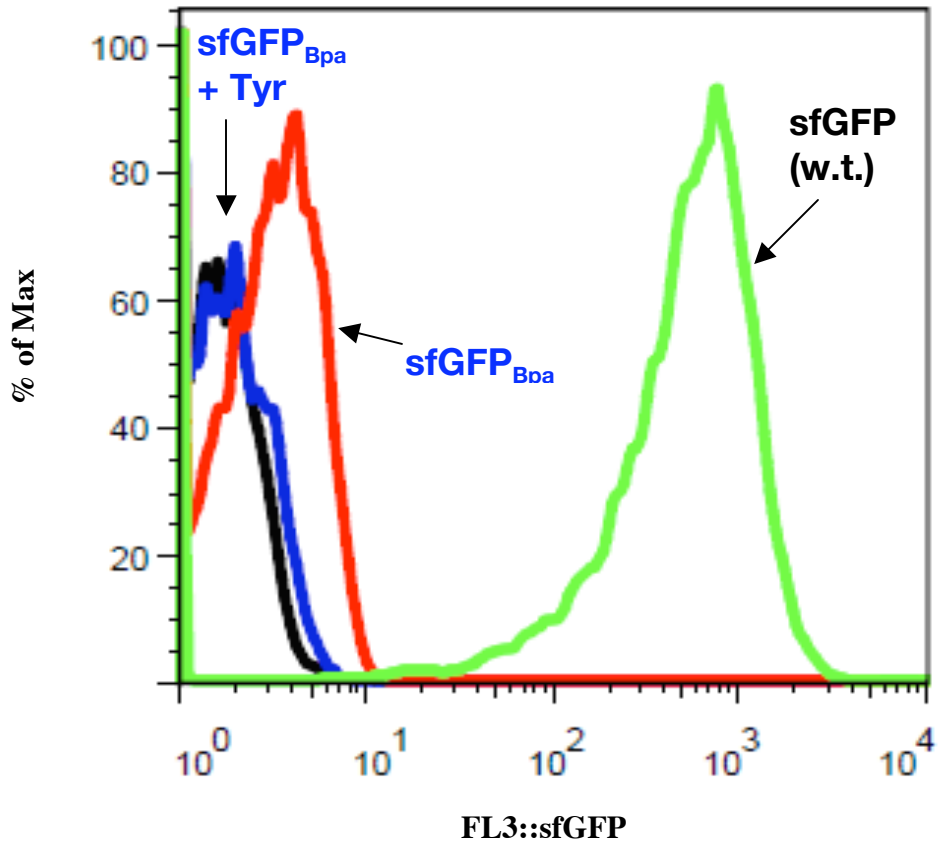


Figure 3. FACS analysis of cells after expression of superfolder GFP in the MRA317 host strain cotransformed with **pIL80** and pDULE-Bpa (red, blue, and black curves) in rich LB media without additional amino acid supplementation (black curve), supplemented with tyrosine (2 mM, blue curve), or supplemented with the unnatural amino acid ***p*Bpa** (2 mM, red curve) at 37 °C. A positive control experiment was also performed with MRA317 cells transformed with plasmid **pIL7** encoding the wild-type sfGFP sequence in rich LB media without analogue supplementation.

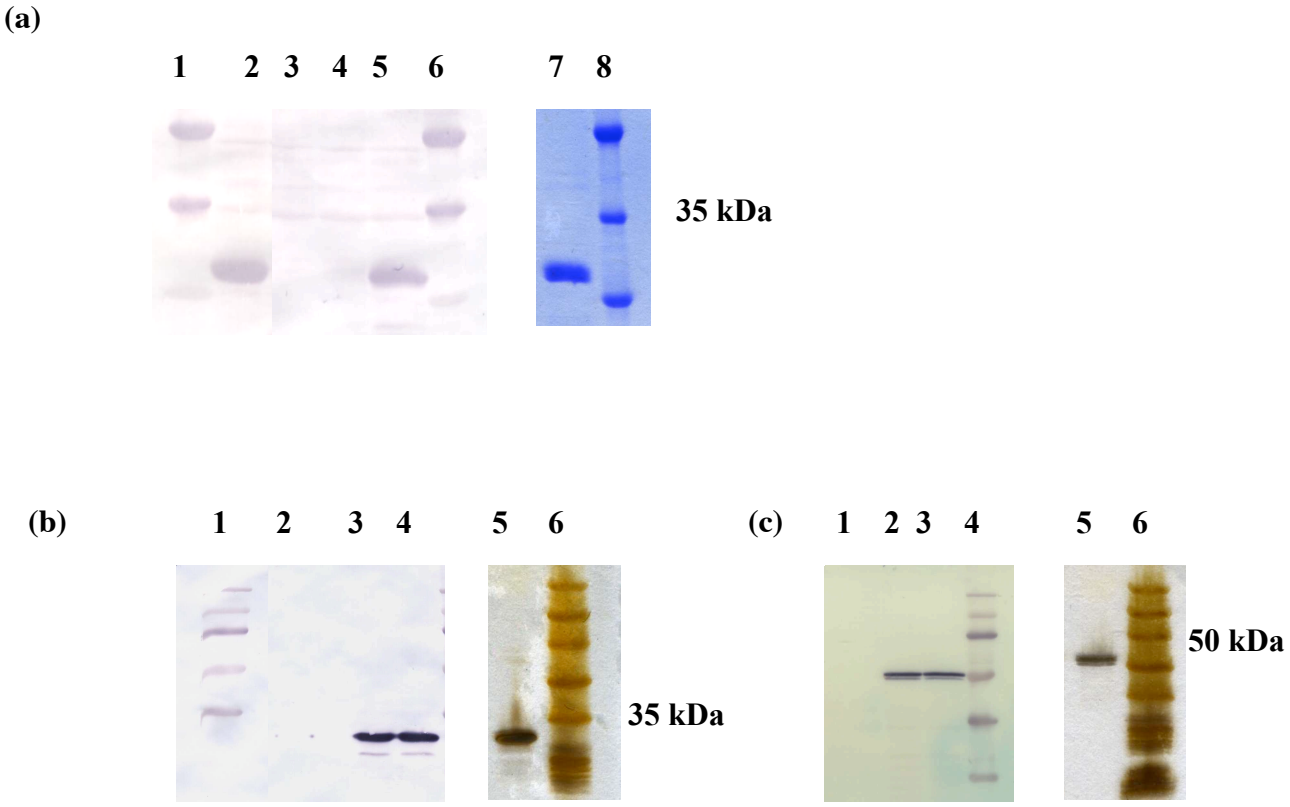


Figure 4. SDS PAGE and Western blot analysis with anti-his•tag monoclonal antibody indicating multi-site incorporation of ***p*Bpa** into a (a) sfGFP variant and the elastin-mimetic polypeptides encoded by (b) **pIL43.5** and (c) **pIL43.11**. SDS PAGE analysis was used to verify presence of purified sfGFP and elastin-mimetic polypeptides with Coomassie and silver staining, respectively. (a) The sfGFP variant was synthesized from *E. coli* strain MRA317 cotransformed with the plasmid encoding the target protein **pIL80** and the plasmid containing the orthogonal tRNA/synthetase pair. Lanes 1,6, 8: Perfect Protein™ Marker, 15-150 kDa (Novagen®, a brand of EMD Chemicals, Inc., Gibbstown, NJ); lane 2: (positive control) expression of wild-type sfGFP from plasmid **pIL7**; lane 3: analogue deficient (negative) control; lane 4: additional tyrosine supplementation (2 mM); lane 5: ***p*Bpa** supplemented (2 mM); lane 7: purified sfGFP variant from cultures supplemented with ***p*Bpa** (2 mM). (b,c) Lanes 1,6: Perfect Protein™ Marker; lane 2: analogue deficient (negative) control; lane 3: ***p*Bpa** supplemented (1 mM); lane 4: ***p*Bpa** supplemented (2 mM); lane 5: purified elastin-mimetic polypeptides from cultures supplemented with ***p*Bpa** (2 mM).

similar expression conditions (~50 mg/L). The reduced yield may be a result of target protein remaining in the insoluble fraction following cell lysis as well as poor binding of the analogue containing variant to the cobalt-charged resin during purification by metal affinity chromatography. Lastly, the identity of the unnatural amino acid within the superfolder-GFP was confirmed by electrospray ionization mass spectrometry (ESI-MS). The ESI-MS of the full-length sfGFP revealed peaks corresponding to the theoretical molecular mass of the wild-type sfGFP protein and the variant with insertion of three *p*Bpa residues (Figure 5 and Table 2).

Encouraged by these results, we were interested in the ability of the modified host strain to support even greater biosynthetic incorporation of the analogue in a single polypeptide. Towards this end, we synthesized an elastin-mimetic DNA cassette elastin-UAG (Scheme1). The DNA monomer encodes a five-pentapeptide repeat sequence where the fourth codon of one pentapeptide is replaced with the stopcodon UAG. DNA cassette concatemerization allowed the synthesis of a library of multimers composed of tandem repeats of the oligopeptide unit. Two concatemers encoding twelve and twenty-one repeats each were isolated and cloned into the expression plasmid. In this way, the degree of concatemerization correlated with the number of UAG codons present in the DNA coding sequence. Thus, the elastin-mimetic polypeptide cloning strategy provided a simple and convenient method for assaying the amber suppression activity of the modified host strain by through selection of concatemer length. The resulting amino acid sequence of the **elastin-UAG$_{12mer}$** and **elastin-UAG$_{21mer}$** polymers encoded by the plasmids **pIL43.5** and **pIL43.11** are shown in Scheme 1. Electrophoretic and western blot analysis of whole-cell lysates derived from these expression cultures indicated that the *p*Bpa analogue supported a detectable level of target protein expression with respect to the negative

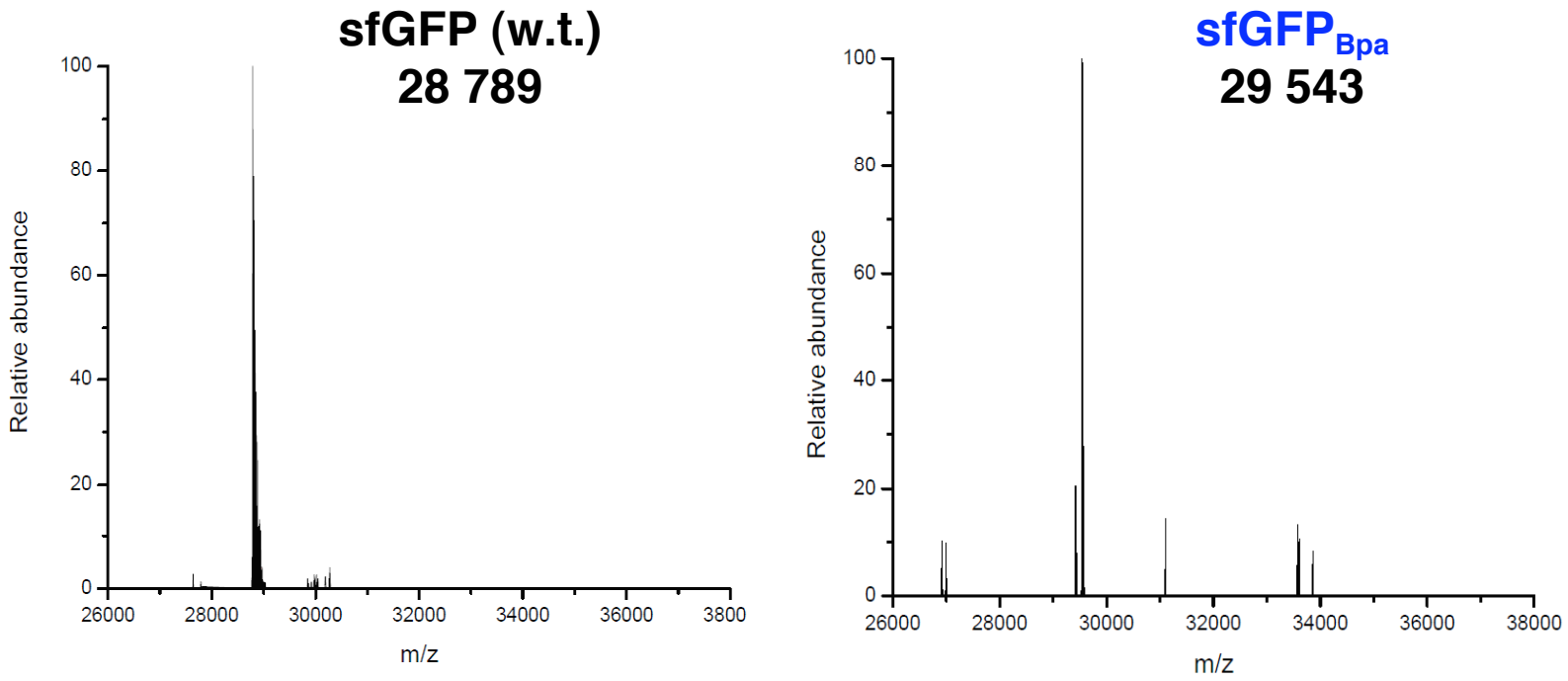


Figure 5. Deconvoluted ESI-MS spectra of the wild-type superfolder GFP (sfGFP) and the sfGFP variant with incorporation of the three **Bpa** residues. The difference in *m/z* of the molecular ions in these spectra (754) agrees exactly with incorporation of three-*p*Bpa residues in the **sfGFP$_{Bpa}$** variant.

analogue supported a detectable level of target protein expression with respect to the negative control culture, which was deficient in the analogue (Figure 4). Protein expression in rich media, lacking the non-canonical amino acid is undetectable by western blot analysis. This result supports the orthogonality of the mutant suppressor tRNA/synthetase pair and shows that the mutant synthetase is unable to charge the suppressor tRNA with the canonical amino acid, tyrosine. The elastin-mimetic polypeptides containing *p*Bpa could be purified to homogeneity from the endogenous proteins of the bacterial host via immobilized metal affinity chromatography (Figure 4). The presence of the decahistidine tag at the C-terminus should ensure purification of the full-length polypeptides selectively. The unoptimized yields of full-length elastin-mimetic polypeptides (Table 1) ranged from approximately 1 mg/L to 2.4 mg/L for fully induced expression cultures in LB media under conventional batch fermentation conditions in a shake-flask culture. The yields are comparable to previous studies investigating *p*Bpa incorporation. Our reported yields are quite respectable, especially considering the high density of amber codons in the mRNA transcript encoding the target elastin-mimetic polypeptide sequences. Thus, expression of the full-length target proteins is supported in the modified expression host, which exhibits increased readthrough of the stopcodon UAG.

The identity of the full-length elastin-mimetic polymer **elastin-Bpa$_{12mer}$** was confirmed via MALDI-TOF mass spectrometry (Table 2 and Figure 6). The molecular ion present in the MALDI-TOF mass spectrum of the full-length elastin-mimetic polymer (12 repeats) was consistent with the calculated values of *m/z*, assuming that the observed ion is singly charged, for incorporation of *p*Bpa in response to the amber codon. Mass spectrometry of the full-length **elastin-Bpa$_{12mer}$** demonstrate that the expression system affords a high level of biosynthetic

Table 2. Expression yield and mass spectrometry data for the sfGFP variants and the elastin-mimetic polypeptides.

Protein	Yield (mg/L)	Calculated *m/z*[a]	Observed *m/z*	|*m/z*| (% error)
sfGFP (w.t.)	56	28 791.3	28 789	2 (0.007)
sfGFP-Bpa	1	29 545.2	29 543	2 (0.007)
elastin-UAG$_{12}$	2.4	28 983.52	28 960.1	23.4 (0.08)
elastin-UAG$_{22}$	1	n.d.	n.d.	

[a] Molar masses for the sfGFP and elastin-mimetic polypeptides were calculated assuming cleavage of the N-terminal formyl-methionine.

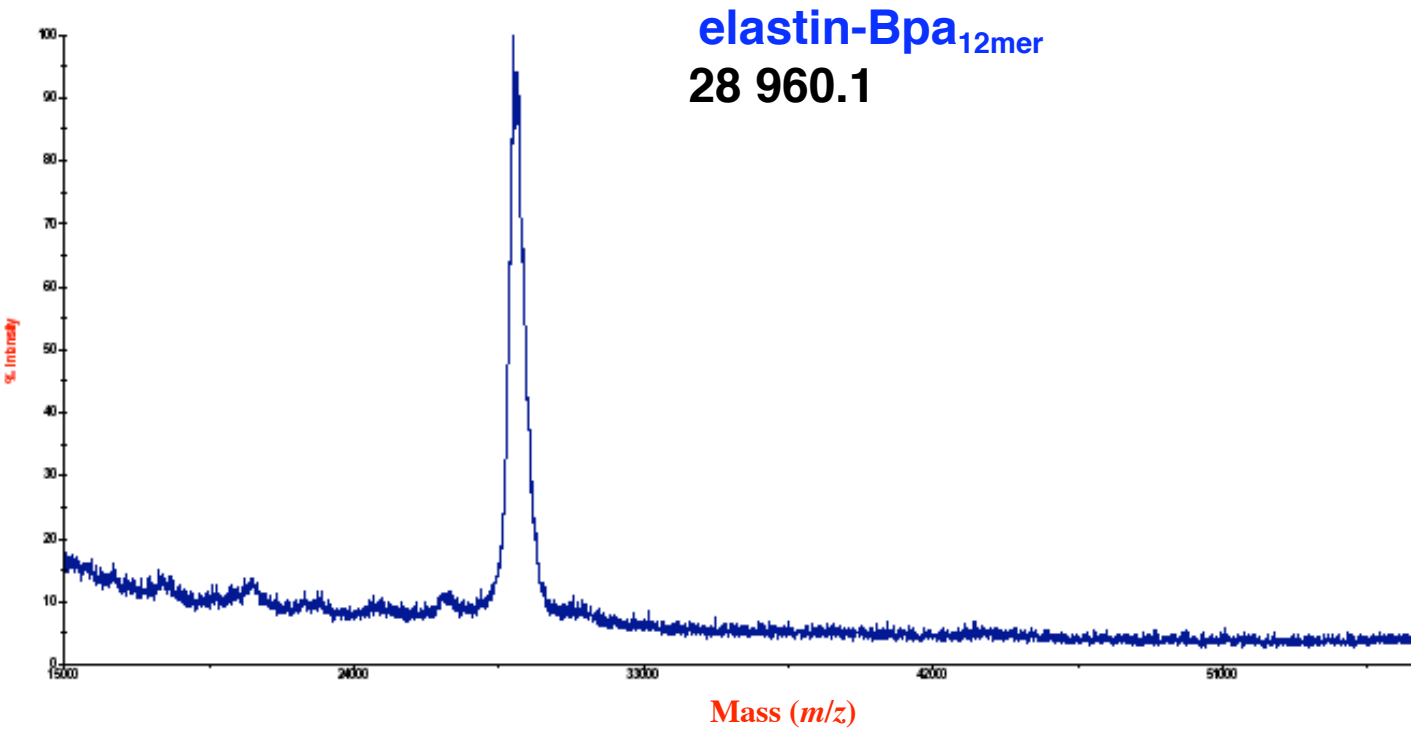


Figure 6. MALDI-TOF mass spectrometry of the **elastin-Bpa$_{12mer}$** polymer isolated from MRA317 cells, with coexpression of plasmids **pIL43.5** and pDULE-Bpa, supplemented with *p*Bpa (2 mM). The *m/z* of the observed molecular ion is consistent with a high level of substitution with the unnatural amino acid.

substitution. Tandem MALDI-TOF/TOF experiments were then performed on thermolysin cleavage products of the *p*Bpa-containing elastin-mimetic polypeptide. Despite the presence of peaks corresponding to a contaminating organic polymer, the spectra support that incorporation of the unnatural amino acid occurred at the fourth position of the pentapeptide repeat where the coding codon was instead the stopcodon UAG (Figure 7). Current efforts are underway to characterize the **elastin-Bpa$_{22mer}$** by mass spectrometry. Preliminary MALDI-TOF mass spectrometry for the full-length elastin-mimetic polypeptide expressed from plasmid **pIL43.11** suggests high level incorporation of the unnatural amino acid using the modified amber suppression method. The peak for the observed molecular ion, however, is broad and additional mass spectroscopy experiments to characterize this polypeptide are desired.

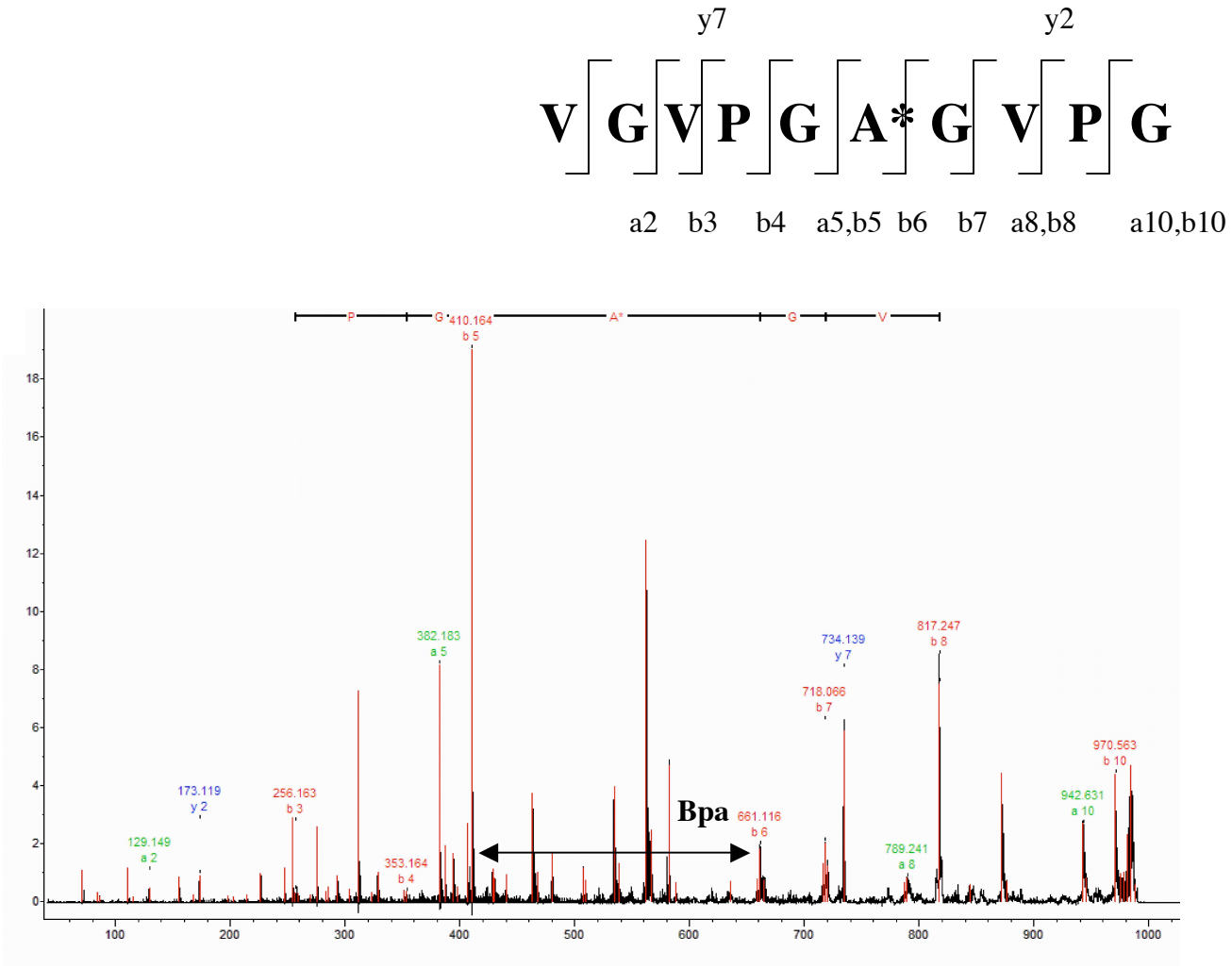


Figure 7. Annotated tandem MALDI-TOF/TOF MS spectrum of proteolytic digestion fragment of **elastin-UAG$_{12mer}$**. The position A* indicates the site of analogue incorporation. The difference between the b5 and b6 ions corresponds to the mass of the unnatural amino acid.

Conclusion

In summary, we have demonstrated the use of a modified *E. coli* expression host that is competent for the incorporation of a structurally dissimilar unnatural amino acid analogue *p*Bpa. The expression system uses an amber suppressor tRNA and mutant synthetase pair from *Methanococcus jannaschii* to charge the unnatural amino acid substrate in response to the codon UAG. The analogue is incorporated site-specifically at multiple sites with a superfolder GFP variant and two elastin-mimetic polypeptides. The multi-site incorporation is possible presumably due to a ribosomal mutation in the host strain, which increases readthrough of the stopcodon UAG. Therefore, multiple amber suppression events may occur within the same mRNA transcript to direct incorporation of *p*Bpa at multiple positions in the growing polypeptide chain. To our knowledge, these results represent a level of biosynthetic incorporation of the unnatural amino acid via amber suppression that is unprecedented. Furthermore, the expression strategy has widespread application for the incorporation of other unnatural amino acids via amber suppressor tRNA/synthetase pairs through cotransformation of the *E. coli* host strain MRA317 with the plasmid encoding the target protein and the pair plasmid. Lastly, DNA cassette concatemerization is an attractive method for assaying amber suppression efficiency, where selection of the DNA concatemer length corresponds to the density of amber codons within the target polypeptide gene.

References

[1] Maskarinec, S. A.; Tirrell, D. A. *Current Opinion in Biotechnology* **2005**, *16*, 422-426.

[2] Van Hest, J. C. M.; Tirrell, D. A. *Chemical Communications* **2001**, *4*, 1897-1904.

[3] Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. *Current Opinion in Chemical Biology* **2010**, *14*, 774-780.

[4] Bilgiçer, B.; Fichera, A.; Kumar, K. *Journal of the American Chemical Society* **2001**, *123*, 4393-4399.

[5] Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790-2796.

[6] Tang, Y.; Tirrell, D. A. *Journal of the American Chemical Society* **2001**, *123*, 11089-11090.

[7] Bilgiçer, B.; Kumar, K. *Tetrahedron* **2002**, *58*, 4105-4112.

[8] Bilgiçer, B.; Xing, X.; Kumar, K. *Journal of the American Chemical Society* **2001**, *123*, 11815-11816.

[9] Bann, J. G.; Pinker, J.; Hultgren, S. J.; Frieden, C. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 709-714.

[10] Seifert, M. H. J.; Ksiazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A. *Journal of the American Chemical Society* **2002**, *124*, 7932-7942.

[11] 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.

[12] Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

[13] Wang, L.; Schultz, P. G. *Angewandte Chemie International Edition* **2005**, *44*, 34-66.

[14] Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nature Chemical Biology* **2008**, *4*, 232-234.

[15] Liu, C. C.; Schultz, P. G. *Nature Biotechnology* **2006**, *24*, 1436-1440.

[16] Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nature Biotechnology* **2007**, *25*, 770-777.

[17] Huang, Y.; Russell, W. K.; Wan, W.; Pai, P.-J.; Russell, D. W.; Liu, W. *Molecular BioSystems* **2010**, *6*, 683-686.

[18] Van Dyke, N.; Murgola, E. J. *Journal of Molecular Biology* **2003**, *330*, 9-13.

[19] Bouakaz, L.; Bouakaz, E.; Murgola, E. J.; Ehrenberg, M.; Sanyal, S. *Journal of Biological Chemistry* **2006**, *281*, 4548-4556.

[20] Rydén, M.; Murphy, J.; Martin, R.; Isaksson, L. A.; Gallant, J. *Journal of Bacteriology* **1986**, *168*, 1066-1069.

[21] Rydén, S. M.; Isaksson, L. A. *Molecular and General Genetics* **1984**, *193*, 38-45.

[22] Moffat, J. G.; Timms, K. M.; Trotman, C. N. A.; Tate, W. P. *Biochimie* **1991**, *73*, 1113-1120.

[23] Dormán, G.; Prestwich, G. D. *Biochemistry* **1994**, *33*, 5661-5673.

[24] Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 11020-11024.

[25] Galardy, R. E.; Craig, L. C.; Printz, M. P. *Nature New Biology* **1973**, *242*, 127-128.

[26] Turro, N. J. *Modern molecular photochemistry*; University Science Books: Sausalito, CA, 1991.

[27] Kauer, J. C.; Erickson-Viitanen, S.; Wolfe, H. R., Jr.; DeGrado, W. F. *Journal of Biological Chemistry* **1986**, *261*, 10695-10700.

[28] Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *Journal of the American Chemical Society* **2000**, *122*, 5010-5011.

[29] Steer, B. A.; Schimmel, P. *Journal of Biological Chemistry* **1999**, *274*, 35601-35606.

[30] Chin, J. W.; Schultz, P. G. *ChemBioChem* **2002**, *3*, 1135-1137.

[31] Farrell, I. S.; Toroney, R.; Hazen, J. L.; Mehl, R. A.; Chin, J. W. *Nature Methods* **2005**, *2*, 377-384.

[32] Sambrook, J.; Russell, D. W. *Molecular cloning: a laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

[33] Payne, S. C.; Patterson, M.; Conticello, V. P. In *Protein Engineering Handbook*; Lutz, S., Bornscheuer, U. T., Eds.; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2009; Vol. 1, p 915-936.

[34] Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nature Biotechnology* **2005**, *24*, 79-88.

[35] Dahlgren, A.; Ryden-Aulin, M. *Biochimie* **2000**, *82*, 683-691.

Chapter 6

Conclusions

The goal of this work was the production of a diverse set of elastin-mimetic polypeptides that exhibit novel biophysical and chemical properties and have the potential to form physically or covalently cross-linked networks. Three broad specific aims were described in addressing this research objective: (1), the design of a cloning strategy that facilitated synthesis of highly repetitive polymers and amphiphilic block copolymers in a manner that is both *modular* and *convergent*; (2), residue-specific incorporation of unnatural amino acid analogues with high levels of biosynthetic substitution that expand the polymer chemical reactivity, which could be exploited under physiologically relevant conditions; and (3), development of a novel expression system for use of an amber suppressor tRNA/synthetase pair to direct site-specific analogue incorporation at multiple sites.

First, the new cloning scheme described and utilized herein permits both DNA cassette concatemerization and directional assembly of the multimers into triblock (**BAB**-type) gene fusions up to 8 kb in length. The method facilitates greater flexibility in the selection of block architecture (size, sequence, order) than previously reported protocols while ensuring seamless cloning between blocks. The cloning scheme was attractive for designing polypeptides where effects of copolymer architecture, specifically the length and density of the hydrophobic block, on high-order copolymer assembly and materials properties could be investigated. The protein hyperexpression protocol allowed for the production of hundreds of milligrams of polypeptide and could be readily optimized for use with large-scale fermentation protocols for biomaterials applications. Lastly, the physical properties of the amphiphilic triblock copolymers were characterized by differential scanning calorimetry (DSC). The polypeptides exhibited inverse temperature transitions above a lower critical solution temperature T_t. The near ambient

temperature maximum of T_t and pH independent thermal transition behavior suggested that temperature-driven assembly of the polypeptides involved selective desolvation and micro-phase separation of the hydrophobic endblocks. The biosynthetic strategy described here has widespread application for the generation of multi-domain polymers composed of an even more diverse set of blocks, such as those derived from sequences of other extracellular matrix proteins (like collagen, laminin), sequences promoting cell-adhesion, growth, migration, or differentiation, sequences signaling protease digestion, etc.

Next, we designed elastin-mimetic polypeptides for unnatural amino acid incorporation at multiple sites using both a residue-specific and site-specific approach. The residue-specific method was employed for co-translational incorporation of the structurally similar analogues of methionine and tyrosine, (*2S*)-2-amino-5-hexenoic acid (Hag) and L-3,4-dihydroxyphenylalanine (DOPA), respectively, to generate novel elastin-mimetic polypeptide derivatives. The polymers were synthesized in relatively high yields in *Escherichia coli* strains auxotrophic for the natural amino acids. The co-translational approach was appropriate for multi-site incorporation of the analogues at a relatively larger number of coding codons (18 to 21 per polypeptide chain). Protein purification also revealed that the cellular accumulation of proteins was improved by increasing activity of the wild type aminoacyl-tRNA synthetase. Taken together, mass spectrometry data, amino acid compositional analysis, and NMR spectroscopy supported high levels of biosynthetic substitution, up to complete replacement of the natural amino acid in the case of Hag. In addition, a diblock copolymer was synthesized where the Hag-containing block comprised the C-terminal domain.

Incorporation of the analogues into the elastin-mimetic polypeptides resulted in temperature-driven assembly, as measured by DSC and UV-visible spectrophotometry, of the polypeptides in a manner analogous to polymers containing the natural amino acid, with two notable exceptions. With regard to the diblock copolymer, hydrophobic aggregation was shifted to a higher temperature and depended significantly on the pH of the buffer, which indicated the role of the more hydrophilic N-terminus domain. Transmission electron microscopy (TEM) of the diblock copolymer following formation of a coacervate at high temperature, as indicated by the temperature-dependent turbidimetry profile, revealed the formation of nanoscale particles. The resulting diblock copolymer has the potential to form both physical and covalent cross-links, which could chemically lock the unique nanoscale morphologies, as well as be further derivatized via the introduced alkene functionality.

The DOPA containing polymer also exhibited unique pH-dependent thermal transition behavior, presumably as a result of the oxidation of the dihydroxy-phenylalanine side chain. At neutral pH, the DOPA polymer appeared to form strong aggregates above the T_t, which may have arisen from cross-linking of semi-quinone or quinone species. While cross-linking of the DOPA residues via use of a chemical oxidizer is a possibility, we demonstrated an alternative strategy that generated metal-ligand cross-links between the oxidized DOPA side chains and Fe^{3+}. The stoichiometry of these Fe^{3+}-catechol complexes was modulated through increasing pH as demonstrated by UV-visible spectrophotometry. Furthermore, gelation of a concentrated solution of the elastin-mimetic polypeptide was observed at high pH in a stoichiometric excess of the metal. The resulting hydrogel could be useful for biomaterial application since its formation occurs under environmentally relevant conditions and metal-ligand cross-links may be

sufficiently strong. Further, the metal-DOPA cross-links have the potential to reform after rupture, which would impart a technologically useful self-healing property to hydrogels.

Lastly, the site-specific incorporation of a structurally dissimilar, photo-activatable amino acid analogue *para*-benzoyl-L-phenylalanine (*p*Bpa) was demonstrated using a novel amber suppression strategy. We have devised a simple approach for introducing the unnatural amino acid analogues at multiple stopcodons through use of a modified expression strain, which is characterized by increased readthrough of the amber codon. The mutant strain permits multiple amber suppression events within a single mRNA transcript, which allows the orthogonal tRNA/synthetase pair to charge the *p*Bpa analogue and add it to the growing polypeptide chain. The elastin-mimetic DNA concatemers were generated with increasing degrees of multimerization to assay the limits of the new amber suppression strategy. Mass spectrometry data of full-length polypeptides, generated from cultures supplemented with *p*Bpa, suggests that at least 12, and possibly 22, amber suppression events are possible during translation of a single mRNA transcript. Optimization of the amber suppression method could permit expansion of the genetic code with multi-site incorporation of structurally dissimilar analogues, representing greater chemical diversity.

The cloning strategy and complementary unnatural amino acid incorporation methods represent promising routes toward the development of synthetic analogues of native elastin. The incorporation of analogues via a residue-specific protocol results in yields that may be optimized for use in materials application, such as cross-linked hydrogels and micelles, or formation of bioconjugates. The novel amber suppression strategy may have significant impact on continued efforts to expand the genetic code by permitting site-specific incorporation at *multiple* sites.

Appendix 1

Sequences of Primers Utilized in Plasmid Sequencing

Primer Name	Oligonucleotide Sequence
M13 Reverse	5’- CAGGAAACAGCTATGAC-3’
M13 (-20) Forward	5’- TGTAAAACGACGGCCAGT-3’
pBAD Forward	5′-ATGCCATAGCATTTTTATCC-3′
pBAD Reverse	5’-GATTTAATCTGTATCAGG-3′
T7 Promoter	5’-TAATACGACTCACTATAGGG-3’
T7 Terminator	5’-GCTAGTTATTGCTCAGCGG-3’
pQE promoter	5’-CC CGAAAAGTGC CACCTG-3’
Type III/IV	5’-CGGATAACAATT TCACACAG-3’
pQE reverse	5’-GTTCTGAGGTCATTACTGG-3’
3’-PROTet Seq	5’-CGCTCGCCGCAGCCGAAC-3’
5’-PROTet Seq	5’-CACATCAGCAGGACGCACTGAC-3’

Appendix 2

Sequences of Interest

metG sequence encoding methionyl-tRNA synthetase (MetRS)

ATGACTCAAGTCGCGAAGAAAATTCTGGTGACGTGCGCACTGCCGTACGCTAACGGCTCAATCCACCTCGGCCAT
ATGCTGGAGCACATCCAGGCTGATGTCTGGGTCCGTTACCAGCGAATGCGCGGCCACGAGGTCAACTTCATCTGC
GCCGACGATGCCCACGGTACACCGATCATGCTGAAAGCTCAGCAGCTTGGTATCACCCCGGAGCAGATGATTGGC
GAAATGAGTCAGGAGCATCAGACTGATTTCGCAGGCTTTAACATCAGCTATGACAACTATCACTCGACGCACAGC
GAAGAGAACCGCCAGTTGTCAGAACTTATCTACTCTCGCCTGAAAGAAAACGGTTTTATTAAAAACCGCACCATC
TCTCAGCTGTACGATCCGGAAAAAGGCATGTTCCTGCCGGACCGTTTTGTGAAAGGCACCTGCCCGAAATGTAAA
TCCCCGGATCAATACGGCGATAACTGCGAAGTCTGCGGCGCGACCTACAGCCCGACTGAACTGATCGAGCCGAAA
TCGGTGGTTTCTGGCGCTACGCCGGTAATGCGTGATTCTGAACACTTCTTCTTTGATCTGCCCTCTTTCAGCGAA
ATGTTGCAGGCATGGACCCGCAGCGGTGCGTTGCAGGAGCAGGTGGCAAATAAAATGCAGGAGTGGTTTGAATCT
GGCCTGCAACAGTGGGATATCTCCCGCGACGCCCCTTACTTCGGTTTTGAAATTCCGAACGCGCCGGGCAAATAT
TTCTACGTCTGGCTGGACGCACCGATTGGCTACATGGGTTCTTTCAAGAATCTGTGCGACAAGCGCGGCGACAGC
GTAAGCTTCGATGAATACTGGAAGAAAGACTCCACCGCCGAGCTGTACCACTTCATCGGTAAAGATATTGTTTAC
TTCCACAGCCTGTTCTGGCCTGCCATGCTGGAAGGCAGCAACTTCCGCAAGCCGTCCAACCTGTTTGTTCATGGC
TATGTGACGGTGAACGGCGCAAAGATGTCCAAGTCTCGCGGCACCTTTATTAAAGCCAGCACCTGGCTGAATCAT
TTTGACGCAGACAGCCTGCGTTACTACTACACTGCGAAACTCTCTTCGCGCATTGATGATATCGATCTCAACCTG
GAAGATTTCGTTCAGCGTGTGAATGCCGATATCGTTAACAAAGTGGTTAACCTGGCCTCCCGTAATGCGGGCTTT
ATCAACAAGCGTTTTGACGGCGTGCTGGCAAGCGAACTGGCTGACCCGCAGTTGTACAAAACCTTCACTGATGCC
GCTGAAGTGATTGGTGAAGCGTGGGAAAGCCGTGAATTTGGTAAAGCCGTGCGCGAAATCATGGCGCTGGCTGAT
CTGGCTAACCGCTATGTCGATGAACAGGCTCCGTGGGTGGTGGCGAAACAGGAAGGCCGCGATGCCGACCTGCAG
GCAATTTGCTCAATGGGCATCAACCTGTTCCGCGTGCTGATGACTTACCTGAAGCCGGTACTGCCGAAACTGACC
GAGCGTGCAGAAGCATTCCTCAATACGGAACTGACCTGGGATGGTATCCAGCAACCGCTGCTGGGCCACAAAGTG
AATCCGTTCAAGGCGCTGTATAACCGCATCGATATGAGGCAGGTTGAAGCACTGGTGGAAGCCTCTAAAGAAGAA
GTAAAAGCCGCTGCCGCGCCGGTAACTGGCCCGCTGGCAGATGATCCGATTCAGGAAACCATCACCTTTGACGAC
TTCGCTAAAGTTGACCTGCGCGTGGCGCTGATTGAAAACGCAGAGTTTGTTGAAGGTTCTGACAAACTGCTGCGC
CTGACGCTGGATCTCGGCGGTGAAAAACGCAATGTCTTCTCCGGTATTCGTTCTGCTTACCCGGATCCGCAGGCA
CTGATTGGTCGTCACACCATTATGGTGGCTAACCTGGCACCACGTAAAATGCGCTTCGGTATCTCTGAAGGCATG
GTGATGGCTGCCGGTCCTGGCGGGAAAGATATTTTCCTGCTAAG
CCCGGATGCCGGTGCTAAACCGGGTCATCAG
GTGAAATAA

tyrS sequence encoding tyrosyl-tRNA synthetase (TyrRS)

ATGGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTA
GCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTACCGCTGACAGCTTGCATTTGGGG
CATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCG
ACGGGTCTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGG
GTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCGAAC
AACTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAG
ATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTAC
AACCTGTTGCAGGGTTATGACTTCGCCTGTCTGAACAAACAGTACGGTGTGGTGCTGCAAATTGGTGGTTCTGAC
CAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTT
CCGCTGATCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCGCAGTCTGGTTGGATCCGAAGAAA
ACCAGCCCGTACAAATTCTACCAGTTCTGGATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTC
ACCTTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAACAGCGGTAAAGCACCGCGCGCCCAG
TATGTACTGGCGGAGCAGGTGACTCGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCAAAACGTATTACCGAA
TGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGGACTTCGAACAGCTGGCGCAGGACGGCGTACCGATG
GTTGAGATGGAAAAGGGCGCAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTTCCCGTGGTCAGGCA
CGTAAAACTATCGCCTCCAATGCCATCACCATTAACGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGAA
GAAGATCGTCTGTTTGGTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGCTGGAAATAA